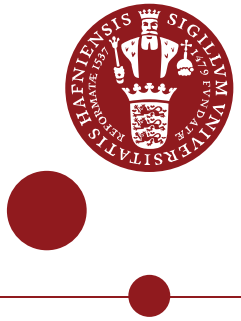
Functional Relations Between Biological Systems

Holt, Louisa Jane

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

UNIVERSITY OF COPENHAGEN
FACULTY OR DEPARTMENT



PhD Thesis

Louisa Holt

Functional Relations Between Biological Systems

Supervisor: Sune Holm, Thor Grünbaum

Submitted on: 19 December 2019

Functional Relations Between Biological Systems

PhD Thesis

By Louisa Holt

Department of Communication

Section of Philosophy

Supervisors: Sune Holm, Thor Grünbaum

Submitted on: 19 December 2019

Word Count: 49, 057 (excl. bib.)

SUMMARY

This dissertation explores the functional nature of biological systems. By adopting and interrogating the organisational approach to functions, I hope to offer fresh insights into biological systems, their functional nature, and how they differ from artifacts.

The introduction gives a brief overview of the history of the machine analogy of the organism and argues that the driving force of this analogy has always been the fact that both organisms and artifacts are functional systems. After exploring a variety of different ways that functions can be understood, I introduce the organisational approach to functions which I both adopt and critique in the articles that comprise this dissertation. I also provide a brief overview of the articles collected in this dissertation, map the links between them, extend some of the arguments within them, and offer some final reflections.

In Article I, I argue that McLaughlin's (2001) organisational account of functions rests upon a value judgement that singles out persistence as the only good and only goal of biological systems. I argue that this value judgement sits at the heart of McLaughlin's account of both teleology and normativity. This raises questions about the extent to which McLaughlin's account can claim to be a naturalistic account. At the same time, the account fails to restrict teleology and normativity to the biological domain.

In Article II, I explore whether the organisational account proposed in Mossio et al. (2009), Saborido et al. (2011), Moreno & Mossio (2015), Mossio & Saborido (2016) and Mossio & Bich (2017) can ascribe apoptotic traits a function. I argue that it cannot. As a solution, I suggest that their organisational account needs to be re-framed as a disjunctive account, following the proposal made by Delancey (2006). This allows for the possibility that biological traits can be both self-maintaining and self-reproducing, they can function for the good of systems they are part of and for other token systems of the same type.

In Article III, I examine the types of teleological relations that arise through various forms of biological interactions and argue against the oft-made claim that all biological teleology is intrinsic. Instead, I argue that many teleological relations in biological systems are extrinsically teleological and are so in view of their cooperative or competitive nature. Analysing the types of extrinsic teleology that arise through symbiotic interactions, I argue that they are fundamentally different from the type of extrinsic teleology possessed by artifacts.

ACKNOWLEDGEMENTS

I would first like to thank my husband, Heine Hansen, for giving me the gift of happy home with a room of my own in which I have pursued this PhD. I would also like to thank my late father, Christopher Holt, who taught me that the value of an education lies in learning how to think, not what to think. And I would like to thank my mother, Joanne Rooney, without whom none of this would be possible.

I am indebted to both of my supervisors, Sune Holm and Thor Grünbaum, whom have given me detailed and insightful feedback on earlier drafts of the articles in this dissertation. I can assure you the final result is much better for their input. I would particularly like to thank Sune for the encouragement, compassion and confidence he has shown to me over the last four years. I have many wonderful memories for which I am grateful to him. And I would like to thank Thor for always pushing me to be the best I can be. I have been extremely lucky to have Thor as a mentor throughout both my undergraduate and graduate education.

I am also greatly indebted to a number of people for their feedback, encouragement and comments on earlier drafts or presentations of the ideas collected in this dissertation. In particular, I would like to thank Maria Serban, Sara Green, Peter McLaughlin, Matteo Mossio, Philippe Huneman and Joeri Witteveen. I would also like to thank everyone who organised and participated in the Ninth Philosophy of Biology and Cognitive Sciences Research Workshop, at the University of the Basque Country; as well as the organisers and participants of the 'Organisms, Living Systems and Processes Conference' held at the University of Exeter in 2017; everyone who participated in the many enjoyable Living Machines? workshops; and attendees of the Manuscript in Progress workshops that run in the philosophy section of the University of Copenhagen. Finally, I would like to thank the Institut d'Histoire et de Philosophie des Sciences et des Techniques at the University of Paris for hosting me as a visiting student, I do hope I make it back sometime.

This PhD project has been funded by the Independent Research Fund Denmark grant number DFF – 4180-00146 as part of the project: *Living Machines? Prospects and Limitations of the Engineering Approach to the Organism.*

PREFACE

I was tasked, in this PhD, to consider the oft-drawn analogy between organism and machine. From this starting point, I have produced three articles which barely mention the words machine, artifact or analogy. Instead, the focus has firmly been on exploring the functional nature of the various relationships that arise between biological systems: between a multicellular organism and the cells that constitute it, between an organism and its lineage and niche, between symbionts, colony members, etc. I have elected to do so within an organizational approach to biological functions.

As I detail in the introduction, the path from the machine analogy of organisms to functions is actually quite a direct one. One of the forces driving the machine-organism analogy is the fact that both systems are functional. I will argue that this fact sits at the heart of the machine analogy. From a distance, this commonality seems to identify machines and organisms as the same kind of systems, they are both functional systems. But closer inspection suggests that machines and organisms are functional for very different reasons and in very different ways (Nicholson 2013). Explicating these differences has been one of the real achievements of the organizational approach to biological functions (McLaughlin 2001, Mossio & Bich 2017). This is one of the reasons I adopt an organizational framework when approaching the analysis of biological functions in this dissertation.

But relationships between biological systems raise a number of thorny issues for the organizational approach (Delancey 2006, Artiga & Martinez 2015, Bich 2019). If these relationships are functional, as biological practice and theory implies, their consideration suggests that the organizational approach is too restrictive and the functional nature of organisms and machines is more similar than claimed. So goes the general arc of the argument made in this dissertation. But before I get started making that argument, I would first like to make some general comments about the connection between the organism-machine analogy and the various relations that arise between biological systems.

The most obvious difference between an organism and a machine is that one is alive, whereas the other is not. If functional status is what brings organisms and machines together, living status is what tears them apart. But what is it that makes one system alive and the other not? According to the organizational approach, it is the fact that living systems are self-maintaining, self-regenerating, self-reproducing and self-determining (Ruiz-Mirazo et al. 2004, 2010). Machines are none of these things.

In this dissertation, I assume that there is some truth to the statement that living systems are self-maintaining. This is one of the key differences between organisms and machines. In addition, it takes seriously a proposal by O'Malley and Dupre (2009) that biological systems are fundamentally collaborative. As I see it, this

points to another important difference between machines and biological systems: biological systems are collaborative, machines are not[1]. But it appears there is something of a tension between these ideas. The basic challenge is this: organizational accounts have claimed that biological systems function for their self-maintenance but the collaborative nature of biological systems suggests that they also function for ends external to themselves.

To tease out this tension, in this dissertation, I consider various types of biological collaborations and how they can be accounted for within an organizational framework. Much of this dissertation is focused on the task of making the challenge organizational accounts face clear. In addition, small steps are made towards resolving the challenge and suggesting pathways for future research that organizational accounts can explore.

[1] Note that this view does not entail the claim that cooperation is an essential property of life. Rather, Dupre & O'Malley argue that biological systems are collaborative. As they put it, 'Collaboration from this point of view covers a range of interactive processes that may include both cooperative and competitive activities' (2009, 2). Thus a collaborative system can be cooperative, though is not necessarily so. It might also be parasitic.

INTRODUCTION

Between organisms and machines

History of the analogy between organism & artifact[2]

For as long as philosophers and scientists have asked the question, 'what is an organism?', there have been those that answered, 'it is a machine'. For as long as this answer has been given, it has been disputed, the disparities between organism and machine being at least equal to their similarities. And yet the idea has remained, constantly reinventing itself to reflect both our evolving understanding of organisms and the technology of the day. Discredited in one form, it re-emerges in another. Discredited in every form, it shapeshifts from ontological stance to heuristic tool or rhetorical device.

The first known comparison between organism and machine is usually attributed to the Greek physician Hippocrates, who likened the human body to a water clock. The latest technology in 400 BC, water clocks were hydraulic devices whose various parts were controlled by a carefully balanced flow of water, in order to tell the time. Reflecting on these technologies, Hippocrates proposed that the human body was made up of four humours which existed as liquids: blood, phlegm, black bile and yellow bile. And like the liquid in the clock, the four humours had to be kept in balance for the body to function (Jouanna 2012).

In the 14th century, mechanical clocks started to appear. The first were driven by a simple system of weights. These were soon replaced by springs and finally in the 17th century, by the spectacular, complex and precise pendulum clocks which would inspire Descartes, La Mettrie and Newton (Riskin 2016). It is hard for us now to imagine how truly marvellous this technology must have seemed. Objects which for all time had been inert and completely dependent on a human operator now sprang into motion, asserting some capacity for self-directed movement. As the philosopher Comenius put it, 'is it not a truly marvellous thing that a machine, a soulless thing, can move in such life-like, continuous, and regular manner? Before clocks were invented would not the existence of such things have seemed as impossible as that trees could talk or stones speak? Yet everyone can see that they exist now' (in McReynolds 1980). For the first time, a link was forged between the inanimate and animate worlds. This link paved the way for a new approach to understanding organisms. It was an approach that did not require vital spirits or mysterious Gods. Just as we explain how the hands of a clock move by referencing only its inanimate parts, so too might we explain how an organism

[2] Many of the ideas as well as earlier drafts of some sections of the text in this section come from Holt (2018)

can walk or talk by referencing interactions between the inanimate matter out of which it is made. In the face of this new technology, the idea that an organism was like a simple water clock quickly seemed outdated. Surely if the organism was a type of machine, it must be like a pendulum clock, belonging to the same family as the spectacular Strasbourg astronomical clock, only ever yet more complex.

The core of the analogy between organism and clock turns on the idea that organisms can be explained in terms of regular, recurring, automatically controlled events among their parts. It positions the body as a purely mechanical thing, which can only be understood by studying it as such. It is hard to overstate the importance of this idea to much of the scientific progress which has occurred since the 17th century. First, the rise of the mechanical view served a huge blow to vital theories of life which had dominated since Aristotle (Brito & Marques 2014). But more importantly for the scientist, it justified adopting engineering methods within biology (*ibid*.). If the organism is a type of machine, then it can be studied like a machine: to understand an organism we need only decompose it into its constituent parts and then identify the functions of those parts.

Hot on the heels of the age of mechanics came the age of heat and steam. In the sixteenth century, the first vacuum and air pumps started to appear. In the seventeenth century, this technology was harnessed by Denis Papin to build the first steam digester, a machine that confined steam in a tight space until a high pressure was created. In Papin's machine, a combination of heat and water was used to create a vacuum, causing atmospheric pressure to build up against a piston, resulting in motive power. Then, the first steam engines appeared, machines which harnessed that pressure to produce mechanical work (Dickinson, 1939). At the same time, and very much hand-in-hand with these technological advancements, the theory of thermodynamics was rapidly developing. This led to a new version of the machine analogy, first popularised by von Bertalanffy (1969) but also elaborated upon by Maturana & Varela in their theory of organisms as *autopoietic machines* (1973/1980). Like the theory of thermodynamics itself, the new analogy was comprised of two parts. The first, based on the first law of thermodynamics, had to do with the conservation of energy. The primary idea was that animals, like machines, must be fed by a constant flow of fuel, which must be burnt, so that the overall energy level remains stable. The second analogy was based on the second law of thermodynamics, the tendency of all systems towards disorder and entropy. In this respect, organisms and machines seem to be very similar. They both seem to act against the general tendency towards disorder. To the contrary, they are both highly organised systems.

Meanwhile, in the 18th century, the first telegraphs appeared; systems which involved the transmission of signals, not just energy. This technology led thinkers to the idea that organisms are not just connected to the world by flows of energy but also by flows of information. By the mid-20th century, this new analogy would find firm expression in the developing field of cybernetics. Popularised largely by Wiener, cybernetics was concerned with systems that process and react to environmental information in order to achieve their goals.

As early as 1944, Wiener was organising meetings that brought together engineers working in communication and control with biologists and medical researchers. In his 1948 manifesto, *Cybernetics: Or Control and Communication in the Animal and the Machine*, Wiener offered a theory of organisms which emphasized messages, codes, information and feedback loops. The new analogy consisted of three parts. First, the machine has sensory organs just like an organism: it has thermometers, light receptors, microphones and all sorts of other tools used to register various signals. Second, the machine has motor functions just like the organism: these are carried out by whatever device the machine uses to create an output, an engine or heating element for instance. Finally, between the sensory and motor organs, the machine has a central regulating system which coordinates incoming messages from the outside world and from its internal organs, and then directs the system to bring about the appropriate reaction.

Finally, in the 1940s Turing invented the digital computer. If cybernetics told us that studying life involves studying the reception, transmission, storage, transformation, and use of information, computation provided a fleshed-out theory for how that information processing actually works. It did so by appealing to algorithms: finite sets of rules that transform data sets in pre-determined ways. It is this appeal that sits behind one of the most prolific versions of the machine analogy in modern science: the claim that DNA is a type of computer program, a genetic code. Building directly on Wiener's conception of the organism, von Neumann presented his vision of a biological computing machine able to reproduce itself by following a set of instructions encoded as a set of genes. The gene was compared to the information tape of a Universal Turing Machine, it was envisioned as encoding a set of instructions for programming an organism (Cobb, 2013). It is an idea that predates Watson & Crick but which remains hugely influential to this day. Just six weeks after they first described its double-helix structure, Watson & Crick would write 'the precise sequence of the bases is the code which carries the genetical information' (1953). This gave rise to the great coding challenge of the 1950s and 1960s, trying to ascertain which letters in the DNA code corresponded to which amino acids. It was a challenge that was finally cracked by Nirenberg & Matthaei, along with their collaborators, in 1967. There is no doubt that framing the challenge of DNA transcription in the familiar terminology of a code helped lead to one of the most important scientific discoveries in the last century.

Considering the history of the machine analogy, it seems plain that the machine analogy has been an invaluable source of inspiration for biologists and philosophers. Each analogy, like each machine, draws our attention to a different aspect of the organism. And while no single machine has proved the perfect model, each analogy has contained a kernel of truth which has been retained, even as the analogy fades away. Point in case, appeals to mechanisms did not cease with appeals to the clock analogy. Likewise, the notions of energy, control systems, feedback loops and programmes all remain entrenched explanatory concepts in the biological sciences. This in itself points to the value the machine analogy has wrought. This is not to say that the machine analogy does not have its problems, it does and excellent work has gone into explicating these problems in recent years (McLaughlin 2001, Lewens 2004, Nicholson 2012, 2013, 2014, 2018). But, despite

these problems, it really does seem inarguable that the analogy also brings with it some value. In the next section, I will argue that this value stems from a fundamental truth which the machine analogy taps into.

The teleological status of organisms and machines

In the previous section, I presented a number of different examples of the machine analogy and emphasized the way in which each analogy, like each machine, was slightly different. But as much as each analogy is different, each analogy is also very similar. They all point to a fundamental fact about machines; machines are teleological systems that were brought into existence by man in order to fulfil a particular purpose. Because of this, each machine is constructed out of specific parts performing specific functions and organised in a specific way that allow the machine to fulfil that purpose. For example, the purpose of the water clock is to tell the time. This is why it was brought into existence and the purpose for which it exists. In order for it to fulfil that purpose, it is constructed out of various parts which each serve a particular function: a container for containing water, a reservoir for supplying water, markings for representing time, and so on. And these parts are all organized in such a way that they enable the clock to tell the time by maintaining the necessary balance.

Now consider the organism. Like a machine, it too is constructed out of various parts which seem to serve particular functions; the function of the heart is to circulate blood, the eyes to see, the ears to hear and so on. And it is this fact which sits at the core of the analogy. Both artifacts and organisms are constituted out of functional parts which are organized in a specific way that enable the system to fulfil its purpose. In the case of a machine, its purposes can be wide and varied but in the case of an organism, its primary purpose is usually taken to be its own survival and reproduction. This implies that it is the assumption of purposeful organization that unites the various versions of the machine analogy. From this, they are differentiated by the particular organizing principle they appeal to in order to explain how those parts enable the system to fulfil its purpose: the water clock is a system organized through a principle of balance, the clock through mechanisms, the steam engine through energy conservation, cybernetic systems through control and information, the computer through a programme.

At this point, it is useful to take a step back and consider the explanatory strategy being pursued here. The general approach takes it for granted that both organisms and machines are purposeful systems and it then explores whether they fulfil their purposes through similar means. There is, however, a fundamental problem with this approach; it assumes that both organisms and machines are purposeful in the same way. But this is

an idea that has been criticized for as long as it has been around and is particularly untenable in the post-Darwinian era[3].

The argument that organisms and machines possess the same type of teleology was first suggested by Plato He proposed an account of teleology explicitly modelled on the teleology of artifacts. He argued that the teleology of artifacts is derived from the intentionality of the artisans who make them. For example, an artisan might have the desire to stay warm. This motivates him to build a house. The house serves the function of providing warmth for the artisan. In this way, the original desire which the artisan possessed explains the apparent purposiveness of the house which is built. Because the purpose or end towards which machines and their parts function as means is determined by or originates in the interests and needs of an external agent, their purpose is said to be extrinsic. Because it is grounded in the desires and beliefs of its maker or user, their purpose is also intentional. Finally, because it appeals to facts about the system's creation, the Platonic model of teleology is usually said to be creationist, it explains the origins of a system's purposiveness. Within Plato's worldview, this model could be non-problematically applied to artifacts brought into existence by an artisan and the natural universe, including all the organisms contained therein, which were brought into creation by Plato's demiurge.

From the get go, it seems inarguable that organisms can have Platonic extrinsic purpose; guide dogs and crops are both examples of this by virtue of the function they serve in society. But Aristotle believed that this is not the only type of purposiveness that organisms possess. In addition to whatever extrinsic ends an organism might serve, Aristotle argued that organisms were intrinsically purposeful, they possessed a purpose which was immanent in the organism itself. It was a purposiveness that was not imposed by an external intentional agent. To the contrary, it arose from within the organism and was usually not intentional at all. By this conception, the purposiveness of organisms is inherently different from the purposiveness of machines. It is the Aristotelian conception of teleology that Galen would adopt in his anatomical studies, arguing that the parts of an organism must be explained by reference to their function in promoting the whole organism (Schiefsky 2007). And it was this Galenic view that would dominate medical thinking right up to the Enlightenment (Allen & Neal 2019).

The rise of mechanism during the Enlightenment signaled a shift away from overtly teleological explanations in favour of explanations which referenced mechanistic interactions of inert matter (French 1994). As noted above, this represented a major step forward in the study of living systems. As the very existence of clocks

[3] For detailed discussions of the distinctions between Platonic and Aristotlian teleology, see Achinstein 1983, Chp. 8, McLaughlin 2001, Chp. 1 and Johansen 2004, Chp. 4. See also Allen & Neal 2019 for a short review in the context of their application to biological systems.

seemed to assert, it might one day be possible to explain the animation of organisms without reference to either vitalistic forces or mysterious Gods. And yet, ironically enough, this shift did not signal the end of references to mysterious Gods or teleology. Instead, in the hands of William Paley (1802), this shift led to the infamous Argument from Design, one of the strongest expressions of the machine analogy of the organism. Comparing the parts of a watch, designed to be so perfectly suited to their tasks, with the remarkably well-adapted features of an organism, Paley reasoned that just as the design of the former points to the existence and skills of its designer, so too must the apparent design in the later point to the existence and skills of its designer. Just as artifacts have Platonic teleology, so too must organisms. By this reasoning, not only are organisms and machines both purposeful systems, but they are both purposeful in exactly the same way.

Paley's *Natural Theol*ogy is famous not only for its persuasive articulation of the Argument from Design but also for its identification of an important type of biological phenomena, biological adaptations. It was a publication that fascinated Darwin during his studies at Cambridge (Darwin 1887, 47). With the publication of *On the Origin of the Species* (1860), he would take the same phenomena, the remarkable adaptivity of organisms, but give them a very different explanation. He would show how adaptive form can result from the differential reproduction of organisms marked by varied heritable adaptations. From the outset, it was quite plain that accepting the theory of evolution meant rejecting the Argument from Design (Ratzsch & Koperski 2019). Organisms were adapted to their environment, true, but this does not necessitate that they are either designed or made by an intelligent creator. In short, both organisms and artifacts might be purposive systems but they are so in very different ways, for very different reasons.

Among Darwin's contemporaries, there was much debate as to whether Darwin had eliminated or merely transformed teleology in biology (Lennox 2010, Sloane 2019). Although this debate continues to this day, most agree that evolutionary explanations are in some sense teleological (e.g. Ayala 1970, 1999, Lennox 1993, cf. Dawkins 1976, 1986); the end for which an adaptation functions is an ineliminable feature of explanation by appeal to natural selection. But these explanations are not teleological in the traditional Platonic sense. Most notably, they are not intentional. They do not reference the intentions of a creator in order to explain how the purposive system came into existence with its particular purposive organization and parts. As such, once again, it has become popular to believe that the purposive nature of organisms and artifacts is fundamentally different in kind.

In light of these developments, a new account of biological teleology was needed. Largely, this work has been carried out within what has come to be known as the functions debate in philosophy of biology. But before turning to discuss the various accounts of functions which have been suggested, it is worth first considering some desiderata for an account of function.

Desiderata for an account of functions

The functions debate in biology is concerned with the need to provide an account of functional explanations. To take the classic example, it is common enough to assert that the function of the heart is to circulate blood. What does it mean to say that the heart has the function of circulating blood? How is this different from saying that the heart causes the circulation of blood? The short response to these questions is that functional explanations are teleological, they suggest that the heart exists and beats *in order to* circulate blood or *for the sake of* circulating blood (Bedau 1992). This is the purpose for which the heart exists and in order to explain why the heart is where it is, doing what it is doing, we must reference this purpose.

Functional explanations are generally accepted to be an ineliminable part of biological explanation (e.g. Ruse 2000, Ratcliffe 2001, cf Matthen 1997). If we were to strip the biological sciences of appeals to functions and their inherent teleology, biologists would be left poorly equipped to answer questions about *why* a trait exists with the specific characteristics it has. All we would be left with would be cause-effect explanations that provide a descriptive picture of *what* the various traits are doing and *how* they interact.

At the same time, however, many have argued that the use of functional explanations in biology is inherently problematic and that most of these problems stem from their teleological nature (e.g. Pittendrigh 1958, Mayr 1974, Cummins 1975). The charges laid are wide and varied. The most problematic are generally taken to be: the positing of Gods, vital forces and intentions were there are none, the entailment of backward causation, and the belief that functions undermine scientific objectivity by depending on normative value judgements which have no place in scientific explanations (Mayr 1992, Allen & Neal 2019).

As a result, any satisfactory account of functions is expected to deal with these problems and, in addition, they must deal with the usual issues that plague any attempt to formulate truth conditions for function ascriptions[4]. They must neither be too inclusive or too exclusive. In combination, these problems have led to a number of specific desiderata by which accounts of functions are standardly measured. I will quickly consider each in turn and then, in the next section will discuss how the various proposed accounts of functions deals with each of these problems. As I will discuss, not all philosophers accept all of these desiderata and not all accounts of functions satisfy them.

[4] Unless otherwise stated, I will assume that philosophers of biology who are engaged with the project of providing an account of biological functions do so with the aim of providing truth conditions for function ascriptions. Some discussion will be had along the way about various positions certain philosophers have taken with regard this objective (e.g. Cummins 1975, Millikan 1989, Neander 1991). Below I discuss these issues in more detail and define my own methodological approach.

Compatibility with naturalism

As outlined above, accounts of teleology and biological functions have a long and checkered association with appeals to all-powerful and unknowable Gods as well as human intentions, none of which seem appropriate within modern biology. For this reason, most accounts of functions are explicitly naturalistic (Allen et al. 1998, 2-3, Allen & Neal 2019). They attempt to provide an account of biological teleology which neither makes nor entails any appeal to mysterious forces or intentionality. In practice, this requires the identification of truth conditions for function ascriptions which are satisfactorily grounded in non-intentional facts about biological systems. This is the first desideratum of an account of functions. By and large, accounts of functions are distinguished in terms of how they satisfy this objective. Different accounts of functions appeal to different facts about biological systems in order to account for their teleological structure.

But the rejection of intentionality and mysterious Gods means that accounts of functions come face-to-face with a number of other problems. Two of the distinctive features of functional explanations is that they seem to entail backward causation between a cause and its effect and, in addition, they are intrinsically normative. Both of these aspects were explained in the Platonic model by appealing to the intentions of the artisan or demiurge. Stripped of this appeal, naturalistic accounts of functions have to face both of these problems afresh.

Direction of causation

One of the most fundamental problems with teleological explanations is that they seem to pervert the direction of causation (Cummins 1975). In a normal causal explanation, the cause explains the effect and it does so because it temporally precedes the effect – it is *antecedent*. But in teleological explanations, the effect is appealed to in order to explain the cause; given that effects occur after causes, this seems to entail some kind of backward causation. Point in case, how can the circulation of blood explain why the heart exists and pumps if the circulation of blood only occurs after the heart has come into existence and pumped?

The Platonic model circumvents this problem by appealing to the intentions of the artisan who created the entity being explained (McLaughlin 2001, Chp.1). By this model, it was the intention of the demiurge who created hearts that they should beat and circulate blood and this explains why hearts exist, beat and circulate blood. Given the intention was antecedent to the heart's existence and beating, whatever other problems the model has, the Platonic model does not entail backward causation. But for philosophers of biology who claim to provide a naturalistic account of biological functions, any satisfactory account must show how it too is compatible with asymmetric direction of causality.

Preserve the function / malfunction distinction

Function ascriptions are not only distinguished from standard explanations because of their teleological nature, but also because they are normative (Hardcastle 2002). This is clearly seen in the fact that function ascriptions go hand-in-hand with the possibility that a trait can fail to fulfil its function. Anything with a function can malfunction. A malfunctioning heart is one that is not doing as it *ought* which we generally take to be *bad*. In this way, the normative status of function ascriptions derives directly from their teleological status. It is because the trait has a purpose that a trait can be better or worse at fulfilling that purpose and it is because that purpose is good that failure to fulfil that purpose is taken to be bad as opposed to merely abnormal.

The Platonic model accounts for the normative dimension of teleological systems by appealing to the value placed on the artifact's purpose by its creator or user and by a value judgement made by the artifacts creator or user regarding how well it fulfils that purpose (Johansen 2004, Chp. 4). Again, the naturalistic rejection of any appeal to intentions or Gods means that naturalistic accounts must provide a new account of biological normativity, which appeals only to non-intentional facts about biological systems in order to account for the function/malfunction distinction.

Preserve the accident / function distinction

The final desideratum for an account of functions is that the truth conditions it identifies are neither too exclusive nor too inclusive (Allen & Neal 2019). This is essential for preserving the distinction between functional traits and non-functional traits, and has proved particularly problematic when considering traits that have accidental utility. For example, noses can *perform* the function of holding up glasses but it seems wrong to say that this *is* their function. As Schlosser puts the challenge, 'explications of functionality should be *appropriate*, i.e., they should match the way uncontroversial function ascriptions are actually made in the sciences and humanities (1998, 319).

In many respects, this desideratum has proved the trickiest for accounts of biological functions. As we will see, nearly all accounts are either too liberal or too conservative; either ascribing functions where there are none or failing to ascribe functions to traits where there should be one. Some accounts even manage to be both too conservative and too liberal at the same time. That this should be the most demanding desiderata should come as no surprise, satisfying it requires getting the truth conditions for function ascriptions right. And this is particularly challenging because it is now well-recognized that biologists use functions to answer two very different types of questions (Ayala 1970, Cummins 1975, Mayr 1993). First, biologists appeal to functions to answer ultimate questions about why a trait exists. For example, *why do we have hearts? In order to circulate blood.* But biologists also appeal to functions in order to answer proximate questions about why a trait does what a trait does. For example, *why does the heart beat? In order to circulate blood.* By one of these criteria, a trait can be functional while, on the other, it might not be; the most notorious examples

being a human appendix which no longer serves any function though once did and the feathers of birds, which have now been co-opted to serve a different function than the one for which they first came to exist. Resulting from this nexus, various accounts of functions have adopted different strategies to account for the non-function/function distinction. Some specifically define functions in relation to either a trait's existence or current activity. Others attempt to provide a unified account of both types of functions while others have argued that we must accept a pluralistic approach to biological functions. In the next section, I will discuss the two basic approaches to biological functions and how each responds to the desiderata demanded of them.

The modern functions debate

I have now identified the broader project which this dissertation falls within; a study of the analogy between organisms and machines. I have also provided reasons why I have elected to approach this issue through a study of the notion of biological function. Given the notion of function sits at the heart of the analogy, if the functional nature of machines and organisms is inherently different as many have claimed, then this seems to present a serious challenge to machine analogy. I have identified some pretty standard and widely-accepted desiderata which an account of functions should satisfy. Now, in this section, I will consider a number of alternative accounts of functions that have been proposed over the years. This will provide us with a theoretical framework which I will appeal to in order to explicate the organisational account of functions in the next section.

The foundations of the modern function debate were laid through the work of Carl Hempel (1959/1965) and Ernst Nagel (1961), who each developed a distinctive mode of conceptualizing functional explanations. According to Hempel, functions are identified with the contribution a trait performs in maintaining a system (1959/1965, 305). According to Nagel, functions are identified with the contribution a trait makes to the performance of the system's characteristic activity (1961, 403).

Both Hempel and Nagel's accounts were developed within the framework of the Deductive-Nomological account of scientific explanation. As a result, their primary goal was to reduce functional explanations to logical arguments (Hempel & Oppenheim 1948). By this approach, a function is conceptualized as a conclusion to an argument which is meant to explain it. One upshot of adopting this approach, was that both Hempel and Nagel's accounts seem to suggest that functional explanations are illegitimate, either the conclusion was invalid or one of the premises unsound (Hempel 1959/1965, Nagel 1961, see also Cummins, 1975). But this did not lead to a rejection of functional explanation in biology. To the contrary, for many the idea that functional explanations were illegitimate seemed so unacceptable that it led a growing consensus that they were ineliminable (McLaughlin 2001, 73).

Despite the flaws wrought by the D-N framework, most modern naturalistic accounts of functions are in some way a direct development of Nagel and Hempel's accounts (McLaughlin 2001, Chp.4-6). Some

explicitly adopt one of their approaches, while others attempt to mix both approaches in order to overcome the weaknesses of each taken in isolation. Given the influence each has had on modern accounts of function, it is worthwhile to quickly discuss the differences between the two accounts. To appreciate these differences, without getting bogged down in the details of the DN-framework they adopt[5], it will be useful to consider how each approach addresses the identified desiderata above.

The most glaring difference between Hempel & Nagel's approaches concern the way they attempt to make functional explanations compatible with asymmetric direction of causation. This corresponds to a difference in the way Hempel and Nagel view the target of functional explanations. According to Hempel, functions explain the existence of a function bearer. They are answers to the question *Why does X exist?* In order to answer this question, an account of functions must look back through time to identify the causes of the function bearer. As such, his account is an example of what has come to be known as a backward-looking or etiological account of functions. By contrast, Nagel assumes that functions explain what a function bearer does. They are answers to the question *What does X do?*[6] In order to answer this question, Nagel must look forward through time to identify the effects of the function bearer. For this reason, Nagel's account is generally taken to be an example of a forward-looking account which is concerned with a trait's disposition.

A second important difference between Nagel and Hempel concerns the way they conceptualise the normative dimension of functional explanations. According to Hempel, functions contribute to the maintenance of a system, they keep the system in working order or maintain it as a going concern (Hempel 1965, 305). The maintenance of the system is taken to be good, therefore functions which contribute to this maintenance are taken to be good. For this reason, Hempel's account is an example of what has become known as a welfare account. By contrast, Nagel claims functions contribute to a characteristic activity of the system (Nagel 1961, 403). This characteristic activity need not, in and of itself, be good for the system. But a trait can be better or worse at contributing to the system's characteristic activity, it can be better or worse at fulfilling its function. In this way, Nagel provides an account of performance norms.

[5] See Cummins (1975) for a detailed critique of both Hempel and Nagel's accounts.

[6] In Cummins' influential critique of Nagel and Hempel's accounts, Cummins claims that both Hempel and Nagel endorse the assumption that 'The point of functional characterization in science is to explain the presence of the item' (1975, 741). According to Cummins, Nagel's account is supposed to explain why the function bearer exists, not what the function bearer does. This is an interpretation of Nagel that has been and remains widely influential (e.g. Allen & Neal 2019, 8). Rejecting this assumption, Cummins goes on to develop an account of functions which explicitly states that function ascriptions can only ever be responses to questions about why a function bearer does what it does, not why it is there. Although Cummins' expression of this idea is undoubtable more explicit than Nagel's, McLaughlin's detailed analysis (2001, Chp.4) shows clearly that many of these ideas were already present within Nagel's (1961) account.

By very different means, both Nagel & Hempel do away with the problem that teleology poses for naturalism. Hempel does away with the problem of teleology by declaring it an invalid and therefore illegitimate form of explanation. By contrast, Nagel does away with the problem of teleology by defining function ascriptions solely in terms of causal dispositions which contribute to the 'characteristic activity' of the system. But while Nagel's approach undoubtedly lays waste to the problem of teleology and any associated incompatibility with naturalism, the cost of this move is that his account becomes too liberal. As he himself recognizes, 'every effect of an item will have to count as one of its functions, if it should turn out that each effect contributes to the maintenance of some goal or other' (Nagel, in McLaughlin 2001, 118). This problem has become the defining problem for dispositional accounts of biological functions, which have adopted Nagel's overall approach.[7]

Dispositional accounts of functions

The canonical dispositional account of function ascriptions was developed by Robert Cummins (1975). The core of Cummins approach lay in an appeal to the process of functional decomposition. As he puts it, 'to ascribe a function to something is to ascribe a capacity to it which is singled out by its role in an analysis of some containing system. When a capacity of a containing system is appropriately explained by analyzing it into a number of other capacities whose programmed exercise yields a manifestation of the analyzed capacity, the analyzing capacities emerge as functions' (1975, 765).

Cummins' account is important for multiple reasons. First and foremost, he makes explicit an assumption in Nagel's account which lay somewhat obscured (see footnote 4 above). He explicitly defines function ascriptions in relation to what a trait does, not why it exists (1975, 749-750). This is the first time that the polysemous nature of the term 'function' is brought into sharp focus. Cummins explicitly rejects appeals to functions in order to answer questions about a trait's existence and instead argues that functions can only ever answer questions about why a trait does what it does (cf. Ayala 1970).

In Cummins' account, we begin functional analysis by identifying the capacity which we are interested in analyzing and then identifying the various roles which different components of the system perform in contributing to that capacity. By this reasoning, the human heart has a function because it contributes to the process of circulation. First, we identify circulation as the process we are interested in and because hearts contribute to that process, hearts are functional.

[7] Given Hempel declared functional explanations an invalid form of explanation, I do not discuss potential counter-examples to his account.

The account has been influential because it seems to correctly capture at least one way in which biologists use the term function and identify those functions, particularly in comparative and functional morphology (Rudwick 1964, Amundson & Lauder 1994). Consider the famous explanation of the heart's function by William Harvey, 'The chief function of the heart is the transmission and pumping of the blood through the arteries to the extremities of the body (1616/1928, 49). This explanation does not seem to make any claim about why the heart exists, its target is what the heart does. Cummins' account has also been hugely influential in psychology, particularly under the influence of the modern mechanists, who have adopted this approach (e.g. Craver 2001).

In addition, it is a particularly robust account; it is completely impervious to counter-examples. Given functions are assigned dependent on the observer's analytical interests, anything can be a function just so long as they align with the observer's analytical interests. This is one of the biggest differences between Nagel and Cummins' account. Nagel limited the spectrum of capacities which functions could contribute to by stipulating that the system must be directively organised towards that capacity (1961, 403). By this reasoning, the system has intrinsic goals which are identified by considering its organization. In contrast, for Cummins, the system was not analysed as having goals itself, rather it is the analyst who has goals; the goal of explaining circulation, metabolism, reproduction, and so on. And it is these goals which determine which traits are or are not functional (Cummins 1975, 761).

On one hand, this is the greatest strength of Cummins' account and has left it moderately unperturbed by the types of counter-examples which have plagued other accounts of biological functions. On the other hand, many have found it somewhat unconvincing as an account of biological functions (Millikan 1989, Neander 1991). By Cummins own admission, any effect could have a function, so long as it contributes to a process which the observer is interested in analyzing: 'Heart activity, for example, keeps the circulatory system from being entirely quiet, and the appendix keeps people vulnerable to appendicitis' (1975, 752). While this may be true, it is simply not true that biologists ascribe functions to heart beats for the sound they make, nor the appendix for their role in causing appendicitis. The first is usually taken as a side effect of the heart's actual function (circulating blood) and the later an example of a non-functional system component. As such, Cummins' account does not preserve the distinction between function and non-function which is generally considered an important desideratum for an account of function. Nor does his account distinguish between functional and malfunctional traits. By his account, if a kidney doesn't contribute to waste removal, it is simply not functional. He has no way of stipulating what a trait should or shouldn't do, only what it does do.

None of this is necessarily problematic for Cummins, his account is not meant to be an account of how biologists use the term function, it is not conceptual analysis. Rather, it is an explication of the practice of functional decomposition (Cummins, 1983). But for many, this result is somewhat unsatisfactory, an account of biological functions is still desirable in order to make sense of the many times that biologists use the term

function which sit outside the bounds of Cummins' account (e.g. Davies 2001, McLaughlin 2001, Millikan 2003). To do this, most other dispositional accounts have attempted to reinstate the divide between function and effect by delimitating the type of capacity which a function can contribute to.

The first avenue which has been explored by proponents of dispositional accounts attempts to link function ascriptions to the contribution which a trait makes to the system's goals (e.g. Boorse 1976). Like Nagel, these accounts are heavily influenced by the pioneers of cybernetics who attempted to provide an account of organisms rooted in cybernetic theory (Wiener 1948, Rosenbleuth et al. 1943, Nagel 1953). As such, they retain two of the ideas which were central to Nagel's account, that Cummins had summarily done away with. First, they appeal to the idea that a system has goals of its own. Second, they reintroduce the idea that these goals can be identified through the system's directive organization. Most notably, if a system experiences internal or external perturbation, the system will tend to compensate for those disturbances so that the goal state is maintained. In the case of biological systems, the goal states are identified with the system's own survival and reproduction. In turn, the existence of this goal-directed organisation is explained by appealing to natural selection (Allen & Neal 2019).

The key problem with the goal-directed approach is that it applies equally to biological and physical equilibrium systems. The classic example of this is a marble which, once thrown into a bowl, will 'pursue' the 'goal' getting to the bottom of the bowl (Bedau 1992, 36). The basic problem is that many equilibrium systems are not goal-directed, but the appeal to directive organization leaves us ill-equipped to distinguish between the two. For this reason, the goal-directed approach is unable to satisfactorily distinguish between functions and nonfunctions. In addition, the approach is ill-equipped to distinguish between malfunctional and functional traits. Just as we can interpret the marble as having the goal of reaching the bottom of the bowl, so too can we interpret a malfunctional trait as having the goal of malfunctioning.

In an attempt to overcome these shortcomings, Bigelow & Pargetter (1987)[8] have proposed the propensity view of biological functions. On this view, 'something has a (biological) function just when it confers a survival enhancing propensity on the creature that possesses it' (1987, 192). By explicitly defining functions in relation to the goal of survival, Bigelow & Pargetter are able to exclude non-biological equilibrium systems from their accounts. Even if the marble is goal-directed, whatever goal it is pursuing is certainly not its own survival. In this way, Bigelow & Pargetter make some effort to preserve the distinction between

[8] It is worth noting that Bigelow was one of the original authors of the cybernetic account of functions (Rosenbleuth et al. 1943) which in turn inspired Nagel (1953, 1961).

function and non-function. In addition, because of the appeal to propensity, their account is able to exclude examples of one-off accidents that happen to contribute to an organism's survival.

Despite this, the account suffers from a major flaw (McLaughlin 2001, 127); the account appeals to propensities to enhance survival instead of actual enhancement of survival. But for any effect a trait makes, it is possible to imagine some situation (even within the natural environment) where that effect might have a propensity to enhance the organism's survival. This means that nearly every effect of a trait has a function. Under the propensity view, there is no way to distinguish between traits that actually contribute to an organism's survival and those that might potentially contribute to an organism's survival, they are both functional. For this reason, despite some effort to the contrary, the propensity view is even more liberal in its ascriptions of functions than either the basic goal-directed approach or Cummins' account.

In sum, dispositional accounts have made a number of important contributions to the functions debate. Through them, it has become clear that functions are appealed to in answer to two very distinct questions, *why does a trait exist* and *why does a trait do what it does*? Dispositional accounts attempt to legitimize function ascriptions in this later sense of the term. It is undeniable that this is an important endeavor. First and foremost, it seems to capture an important aspect of biological practice; biologists really do care about what a trait does in the current system and how that trait contributes to the functioning of the system as a whole. Many ascriptions of functions simply don't seem connected to questions about a trait's origins. But in the end, all of the proposed dispositional accounts suffer from a fundamental weakness, they are too liberal. They seem to ascribe functions in all sorts of counter-intuitive situations; appendixes are functional in the process of appendicitis, marbles are functional in reaching the bottom of bowls. As such, dispositional accounts are ill-equipped to preserve with the distinction between malfunction and function or accident and function. To date, Cummins is one of the few who have been satisfied with this result, but Cummins is not concerned with accounting for what biologists mean when they ascribe functions to biological traits, his is not a project of conceptual analysis. For all of these reasons, many philosophers of biology who are engaged with conceptual analysis have instead adopted Hempel's approach to understanding function ascriptions, they have adopted an etiological approach to biological functions.

Etiological accounts of functions

Etiological accounts define function ascriptions in terms of their causal history, and they claim that functions say something about a trait's origins, why a trait exists (Ayala 1970). Over the years, numerous different versions of etiological accounts have been suggested (e.g. Wimsatt 1972; Wright 1973, 1976; Brandon 1981; Ruse 1982; Milikan 1984, 1989; Neander 1991; Godfrey-Smith 1994; Garson 2016). Although there are important differences between these accounts, in common, they all develop Hempel's original proposal by appealing to some kind of feedback mechanism (McLaughlin 2001, Chp.5). In this section, I will concentrate on three important versions of the etiological approach.

The first major milestone in the development of the etiological approach is the account of functions developed by Wright. He proposed a simple account of functions by which, to say that 'the function of X is Z means:

X is there because it does Z,
Z is a consequence (or result) of X's being there' (1973, 161).

The first condition says that a trait is there because of the function it performs and the second condition says that the function occurs as a result of the trait being there. This creates a causal loop in which the existence of a trait is causally explained by the effect which it produces (Millikan 1989). But this seems to commit Wright to backward causation. For this reason, as I detail below, all later etiological accounts explicitly introduce a distinction between traits and tokens (e.g. Millikan 1989, Neander 1991, Godfrey-Smith 1994).[9] The second major issue with Wright's account is that it is too liberal, it ascribes functions to all kinds of traits and systems which we do not intuitively think of as functional (Boorse 1976). According to Wright, any trait which is maintained through some kind of feedback mechanism is functional. But this seems to imply that many simple physical systems are functional. The counter example given is usually a stick which gets stuck against a rock in a stream creating a vortex, which in turn keeps the stick stuck against the rock. The stick is there because it creates a vortex and the vortex is a consequence of the stick's being there. Going against most intuitions, Wright's account seems to imply that the stick is functional. That is, it fails to properly preserve the distinction between accidents and functions. Sticks in vortexes are simply not the type of things we do (or should) ascribe functions to (Boorse 1976, Bedau 1992, McLaughlin 2001, Chp.5).

To overcome the shortcomings of Wright's account, a number of philosophers have either implicitly (Millikan 1984, 1989) or explicitly (Neander 1991, Godfrey-Smith 1994) anchored Wright's account in the theory of natural selection. [10] On this approach, we ascribe functions to traits based on their history: the performance of the function by past tokens of the trait must have contributed to the production of the current token trait. The approach has a number of advantages. First, it makes clear the token/type distinction which was, at best, vague and obscure in Wright's own formulation. In this revision, a token trait has a function Y by virtue of the fact that it is a reproduction of a token which also had the function Y, and that the

[9] It is worth noting that this distinction does seem to be implicitly entailed within many of the examples Wright references (e.g. see Wright's discussion of first-generation trait in 1973/1998, 74; see also Nagel's 1977 interpretation of Wright, as well as McLaughlin 2001, Chp.5).

[10] Most authors describe Millikan's account as a development of Wrights (e.g. Allen et al 1998, 12; McLaughlin 2001, 102; Delancey 2006, 72; Mossio & Moreno 2015, 68), though Millikan explicitly rejects that there is any overlap between her account and Wright (Millikan 1989/1998, 311).

performance of the function Y by the reproducing token contributed to the production of the reproduced token. This ensures that the account does not entail backward causation, it does not entail a causal loop between a trait's own existence and its effects. Instead, it draws a causal connection between the effects of previous tokens and the existence of the current token.

In addition, the account can distinguish between the 'proper functions' of biological traits and any accidental utility which biological traits might serve. Noses were selected for many things, but holding up glasses is not one of them and so it is not the function of noses to hold up glasses, even if it is advantageous for a vision-impaired person that they do so. Likewise, the heart does not have the function of producing heart beat sounds, it has the function of circulating blood. So these approaches go some way to preserving the distinction between accident and function. Finally, the account also provides a reasonable basis for making judgements about what a trait *should* do; they should perform the function for which they were selected. In this way, the etiological approach provides a robust account of the nature of biological normativity and the function/malfunction distinction. A malfunctioning liver is not doing what it was selected for.

These strengths speak in favour of the selected-effects approach to functions. That said, the account has been found vulnerable to numerous counterexamples. First, although it is not vulnerable to simple Boorse-style situations, Bedau has shown (1992) that the approach still seems to ascribe functions to a number of more complex physical systems. For example, clay forming crystals possess particular traits which cause them replicate. These replications will also possess this trait and do so in lieu of the effect that the trait had in previous token crystals. This means that many token crystals with that trait exist because previous token crystals possessed that trait. But, like sticks in streams, there are few philosophers who want to ascribe functions to crystals and no biologists who actually do (e.g. McLaughlin 2001, Delancey 2006). So although the approach goes some way to delimiting between functional and non-functional traits, it does not quite seem to go the whole way. It ascribes functional status to a range of physical systems which nobody ever actually ascribes functions to.

A similar problem arises in the biological domain when one considers how the etiological account treats vestigial traits. Again, the account seems forced to ascribe functional status to a kind of trait which nobody ever actually describes in functional terms. Vestigial traits are those traits which spread through natural selection because, at that point, they conferred some kind of adaptive value to an organism. As such, according to etiological accounts, such traits are functional. But due to changes in selective pressures, vestigial traits no longer have any adaptive value for the organism that possesses them. In fact, some, like the human appendix, are downright maladaptive. Because they have no current utility for the organism they are part of, few biologists would ever talk about them as having functions. They *had* a function, true, but it does not seem right to say that they still *have* a function. But the etiological approaches proposed by Millikan and Neander are unable to distinguish between traits that had and have functions. The etiological account must

continue to ascribe them functions until they eventually become replaced by some advantageous variation. In this way, the etiological approach is too liberal. It counter-intuitively ascribes functional status to a whole class of traits which few biologists would ever refer to in functional terms.

In an attempt to overcome this problem, Godfrey-Smith (1994, see also Griffiths 1993 for a similar proposal) has proposed a development of the etiological account, which he calls 'the modern history theory of functions'. By this approach, it is not the entirety of the trait's selective history which determines its current function, but only its recent history. However, as various authors have pointed out, (e.g. Schlosser 1998, 304; McLaughlin 2001, Chp.5), while this goes some way to limiting the number of problem cases that might arise, it does not actually offer a principled solution to the problem. It only makes the problem much rarer.

Not only is the etiological approach too liberal, ascribing functions were there do not appear to be any, it also seems to be too conservative, failing to ascribe functions to traits which intuitively seem functional. Specifically, it is unable to ascribe functions to new traits which do seem to have some kind of utility for the organism they are part of, but which have not been around for long enough for natural selection to have operated over them. Where vestigial traits had the right kind of history, new traits have the wrong kind of history; they have no history. Because they don't have the right history, the etiological account cannot ascribe them functional status. This was a problem which was explicitly recognized by Wright (1973) and has plagued etiological approaches ever since.

To make the problem clear, it is often raised in the context of Davidson's notorious swampman (Davidson 1987, Millikan 1989, Neander 1991, Mitchell 1995), a physically identical double of some individual (I will call him John) who emerges instantaneously through a freak accident. Physically, John and swampman are identical. They are constituted out of identical physical matter organised in exactly the same way. If a comparative anatomist opened them up, they would find two 'hearts' in the same place, looking the same and seemingly doing the same thing. By most intuitions, if John's heart has the function of circulating blood, so too does swampman's 'heart'. But, according to the etiological approach, swampman's 'heart' cannot have the function of circulating blood because his heart is not a reproduction of an antecedent heart that circulated blood. It has no selective history. As Millikan freely admits, 'Such a double has no proper functions because its history is not right. It is not a reproduction of anything, nor has it been produced by anything having proper functions' (1989, 292).

Whatever import you attach to intuitions about a swampman whose appearance are so unlikely as to be near impossible, the problem remains for etiological accounts, from the perspective of the etiological approach, there is no principled difference between the emergence of a new trait through random mutation and the emergence of a swampman. As Delancey puts it, 'the difference between what actually happens in evolution, and occurrence of an *ex nihlo* organism, is a matter of degree, not of kind' (2006, 76).

There are a number of different ways that proponents of etiological accounts have responded to these counter-examples. Millikan has eschewed the importance of such counter-examples on methodological grounds (1989, 296). Explicitly rejecting the philosophical objective of conceptual analysis, Millikan has characterized her account as providing a theoretical definition of biological functions. Given she is not trying to identify truth conditions which make sense of the way biologists actually use functional explanations, counter-examples which are underpinned by philosophers' intuitions or analysis of how biologists apply functions don't matter. Neander, on the other hand, is explicitly committed to providing a conceptual analysis of the way biologists appeal to functions (1991). To square her account with these counter-examples, she argues against the intuition that biologists could ascribe functions to a trait of the swampman (1991, 180). Both of these responses have been found wanting by critics (e.g. McLaughlin 2001, Chp.5; Delancey 2007, 74).

It is often-claimed, especially by proponents of organizational accounts, that there is a common problem with the etiological approach which underpins all of these various counter-examples: they ignore the current utility which the trait has for the organism it is part of. As Delancey puts it, 'biologists and the biological sciences do not posit teleofunctions as historical, external entities, but rather as activities, generally of internal structures, that have current utility to the organism' (2006, 73; see also McLaughlin 2001, 296; Christensen & Bickhard 2002, 11-15; Moreno & Mossio 2015, 68-69). This is both the strength and weakness of the etiological account. It is a strength because it provides a particularly robust theory of biological normativity. But at the same time, it is its Achilles heel; it makes vestigial traits functional despite having no utility for the organism they are part of and it makes new traits not functional despite the utility they serve. In the next section, we will see how the organizational account attempts to draw on the strengths of both etiological and dispositional accounts. In doing so, it manages to avoid many of the counter-examples which have plagued those accounts but it seems to open itself up to a whole new category of counter-examples.

Organisational accounts

Over the last twenty years, a number of authors have independently proposed a new organisational approach to biological functions (Schlosser 1998, Collier 2000, McLaughlin 2001, Christensen & Bickhard 2002, Edin, 2008, Delancey 2006, Mossio et al. 2009, Saborido et al. 2011, Moreno & Mossio 2015, Mossio & Saborido 2016, Mossio & Bich 2017). It is this account of functions that I discuss in the articles collected in this dissertation. There are a number of reasons I have elected to focus on this approach to function ascriptions, rather than any other.

First, it is a new approach so many of the kinks are yet to be worked out, its strengths and weaknesses yet to be fully understood. As we will see, it is also quite an ambitious account; it attempts to account for both the

nature of biological functions and the nature of biological organization. For both these reasons, there are many things worth saying about the organisational approach. Second, an increasing number of the philosophical discussions about the machine analogy of the organism have either explicitly referenced the organisational approach or have quite obviously been inspired by it (e.g. Nicholson 2013, 2014, 2018, Holm 2014). In turn, many proponents of organisational accounts have themselves weighed in on the debate over the machine analogy of the organism (e.g. McLaughlin 2001, Mosiso & Bich 2017, Militello & Moreno 2018). I take aim at a number of the claims made by these authors in Article III of this dissertation. Third and finally, it may not be perfect and it is probably not the only account of functions to do so, but I believe the organisational approach taps into a fundamental truth about the nature of biological functions and the way biologists ascribe them. It manages to capture many of the strengths of the etiological account but improves it by drawing a connection between why a trait exists and what it actually does in the system it is part of. As I see it, this not only results in a more satisfactory theoretical account, but it also reflects the actual way biologists ascribe functions.

Consider the way biologists ascribe functions in the study of symbiosis. On the one hand, most biologists studying symbiosis explicitly endorse a selected-effects etiological account of biological functions. As Monds & O'Toole sum up the standard view, ''Purpose' is an evolutionary term in kind because to say we know the purpose of something indicates that we know the specific pressures that selected for that specific trait' (2008, 75). On the other hand, when Monds & O'Toole go on to describe how biologists identify functional traits in symbiotic interactions, it is quite clear that, in practice, they appeal to quite different criteria. Consider their discussion of the voids that appear in biofilm communities. The hypothesized function of these voids is to channel nutrients into and waste out of the depths of the biofilm. In order to establish this as the function, Monds & O'Toole argue that 'biofilm researchers must demonstrate that a change in channel structure that affects flow of nutrients and/or waste products has negative consequences for the growth and/or viability of a biofilm. Structure must be linked to biological consequence to enable appropriate inference of function (*ibid.*). The point, which has been made by many proponents of the organisational account, is that before a biologist can even think about inferring a selected-effects function for a trait, they must first establish exactly what function the trait is serving now, in its present organization. It is this sense of the term function which dispositional accounts and now organisational accounts attempt to capture. But the organisational approach takes one important further step than dispositional accounts, they attempt to show that what a trait does is inextricably linked to why it exists.

The gist of the organisational approach identifies functions with the role a trait plays in maintaining an organised system, of which it is a vital part and on which its existence depends. As Delancey sums it up, 'the basic insight is simple but powerful. What makes a structure or event S have a teleofunction F is that it plays a role in a complex system, such as an organism, that in turn allows and supports S, including S doing F' (2006, 82). Take the example of the heart. The heart beats which causes blood to circulate throughout the

body. In turn, blood circulation helps to repair and regenerate the body's cells by transporting nutrients to and waste from those cells, maintaining pH and so on. Included in this are the very cells which make up the heart. So, the heart contributes to maintaining the organization of which it is a part and, in doing so, contributes to maintaining itself. According to the organisational approach, it is this connection between a trait's activity and its own existence that naturalizes function ascriptions. Functions are identified with those effects a trait has that contribute to the persistence of the trait by contributing to the persistence of the system it is part of. On this basis, when we pose the question, *why does a heart exist?* We can legitimately respond *Because it circulates blood*.

It is worth noting that in some organisational accounts, functional traits are identified with structures, for example, the function of the *heart* is to circulate blood (McLaughlin 2001, Delancey 2006, Mossio et al. 2009). In others, functional traits are identified as processes or events, for example, the function of the *heart beat* is to circulate blood (Collier 2000, Christensen & Bickhard 2002, Edin 2008) and in others again, functional traits are identified as states of structures, for example, the function of the *beating heart* is to circulate blood (Schlosser 1998). By and large, in this dissertation, I speak of the function of structures unless it otherwise makes a difference.

From the outset, it is clear that the organizational approach has a strong connection to pre-existing etiological accounts. First, like etiological accounts, the organizational approach is explicitly committed to providing an account of *why a trait exists.* As McLaughlin puts it, this is what functions explains (2001, 209; see also Mossio et al. 2009, 813). In addition, like etiological accounts, the organizational approach naturalises this claim by appealing to a 'feedback' mechanism through which the effects of a trait lead to its own maintenance (e.g. McLaughlin 2001, Christensen & Bickhard 2002, 18, Edin 2008, 203, Moreno & Mossio 2015, 31). To see exactly how similar these approaches are, consider Bedau's (1992) analysis of Wright's account:

> 'The heart pumps blood in order to circulate it. According to the etiological analysis, this means that the heart pumps blood because doing so contributes to circulating it. How are we to understand this? Well, the heart pumps blood only because the creature possessing it is alive, and the creature is alive only because blood is being circulated throughout its body. If the heart stopped circulating the blood, the creature would die and its heart would stop pumping. In this way, part of

> the explanation of why the heart pumps at all is that the heart's pumping contributes to circulating the blood' (786).

This account of the heart's function seems to be just as representative of the general organisational approach as it is of Wright's etiological account.[11] Indeed, as Delancey (2006) and Mossio et al. (2009) explicitly acknowledge, the organizational approach is a development of Wright's account but with one important addition. It stipulates that functions only arise within a specific type of organization; it is because they are embedded in this organization that we can explain the existence of a trait by referring to its effects. By this interpretation, both the etiological and organizational approaches develop Wright's account by appealing to a feedback mechanism, but they differ with respect to which feedback mechanism they identify. Millikan, Neander and Godfrey-Smith all build upon Wright's account by appealing to natural selection, organisational accounts all build upon Wright's account by appealing to the unique organization of biological systems.

The claim that biological organization can naturalise function ascriptions faces the organisational account with an immediate challenge. When etiological accounts reference natural selection, they ground their account of functions in a well-established biological theory which supports the claim that traits re-produce themselves over generations in lieu of the effects they have. But what is this theory of biological organization which the organisational account of functions points to and how does it naturalise function ascriptions?

The organisational account of biological systems

What is needed is a well-defined account of biological organization that the organizational account of functions can appeal to. For this reason, the organisational account of functions goes hand-in-hand with the organisational account of biological systems. By far, the most detailed and discussed of these is the account of closure which has been developed in a collaboration between Moreno, Ruiz-Mirazo, Mossio, Arnellos, Barandiaran, Saborido and Etxeberria (summarized and synthesized in Moreno & Mossio 2015). But similar accounts have also been proposed by Collier (2000), McLaughlin (2001), Edin (2008) and Christensen &

[11] E.g. Consider how similar Bedau's analysis of Wright's account of the function of the heart is to the following accounts of the heart's function offered by proponents of the organizational approach: 'heartbeats are intra-generationally functional, because (1) heartbeats are necessary to pump the blood and (2) this ensures that the organs of the body including the heart are kept working' (Schlosser 1998, 326); 'The heart has the function of pumping blood since (C1) pumping blood contributes to the maintenance of the organism by allowing blood to circulate, which in turn enables the transport of nutrients to and waste away from cells, the stabilisation of body temperature and pH, and so on. At the same time, (C2) the heart is maintained under various constraints exerted by the organism, whose overall integrity is required for the ongoing existence of the heart itself. Lastly (C3), the organism realises closure, since it is constituted by a set of mutually dependent structures acting as constraints, which, by contributing in different ways to the maintenance of the organisation, collectively realise self-maintenance (Moreno & Mossio 2015, 73)

Bickhard (2002). In common, all of these accounts point to two features of biological organization which is claimed to naturalize function ascriptions without entailing backward causation.

The first feature they point to concerns the causal structure of biological systems. In one-way or another, all organizational accounts stress the 'closed', 'circular', 'looping', or 'quasi-holistic' causal structure of biological systems which allows biological systems to maintain themselves through time. [12][13] This emphasizes four features of biological organisation. First, biological systems are constituted out of mutually-dependent traits. As Christensen & Bickhard put it, 'each of the processes that form part of the system requires outputs from other processes in the system to function, and in turn contributes to the requirements of other processes' (2002, 4). In particular, it is claimed that every trait in a biological system must be involved in at least two direct dependency relations (Christensen & Bickhard 2002, 10; Moreno & Mossio 2015, 18-24). In one of these relations, another trait in the system must be dependent on it. In the other, it must depend on another trait in the system. If every trait in the system is involved in (at least) two *direct* dependency relations, then this makes every trait in the system *indirectly* dependent on every trait in the system. In this way, all of the traits in the system are mutually-dependent on each other. They could neither exist nor perform their function outside of the system. Second, it is usually emphasized that biological systems are constituted by functionally differentiated traits (Mossio et al. 2009, 10). This stipulates that each component of a biological system must make a differentiated contribution to the maintenance of the system, they cannot all be making the same contribution. Third, it is usually emphasized that biological systems are able to compensate for internal and external perturbations through various control mechanisms. These mechanisms include both the feedback loops appealed to in in cybernetic accounts and other forms of regulative control through which the system can shift from one type of organization to another (Christensen & Bickhard 2002, 18; Edin 2008, 206; Moreno & Mossio 2015, 31-38).[14] Fourth and finally, it is emphasized that biological systems create external boundaries which allow them, to some extent, to determine how they interact with the external environment. For example, through the formation of a cell membrane, a cell can control which materials comes into and out of the cell, allowing them to maintain the optimal level of nutrients inside of the

[12] E.g. 'The only substantial assumption underlying the physiological practice of attributing functions… is that they are biological systems. That is to say, it is assumed that they are either part of or themselves systems that have the capability of *self-maintenance*, a capability that, as far as is known, can only be manifested if the system contains feedback control systems' … (Edin, 2008: 217) and 'when feedback occurs, a causal chain forms a *loop* or *circle'* (Edin, 2008: 207); 'Closing these open loops at the ends require that products of the process are needed earlier as inputs.' (Collier, 2000: 282); 'appealing to a causal loop at work in the organisation of the system, provides an argument for naturalising both the teleology and normativity of functions' (Moreno & Mossio, 2015: 73); 'A kind of historically stretched quasiholistic causal relation similar to that displayed by natural selection is encountered here within one generation' (McLaughlin, 2001: 210).

[13] It is worth noting that there is some debate as to whether the idea also entails downward or inter-level causation (see Moreno & Mossio 2015, chp.3 for a full discussion, as well as McLaughlin 2001, p.173)

[14] See also the discussion on different control mechanisms in Article III of this dissertation: 'The extrinsic purpose of symbiotic organisation'.

cell (Ruiz-Mirazo & Moreno 2012, 33). Adopting the terminology proposed in Moreno & Mossio (2015), who in turn follow Varela (1979), I will hereafter refer to systems which possess all four of these characteristics as closed systems.

The second feature organisational accounts point to concerns the identity conditions of biological systems. It is an additional core tenant of organizational accounts that biological systems have unique identity conditions which distinguish them from non-living systems. As McLaughlin sums it up: 'A system is self-reproducing in this technical sense only if its identity over time is constituted by the activity of constantly replacing its parts. It stays the same by rebuilding or reproducing itself' (McLaughlin 2001, 201). The basic point is that we do not and cannot identify biological traits (or systems) with the matter they are made out of, for this matter is constantly changing. Often, this point is made by appealing to thermodynamics (e.g. Moreno & Mossio 2015, 7). Biological systems are thermodynamic systems that exist in far from equilibrium conditions. As such, the matter out of which they are made constantly decays and must be replaced if the whole organization is to persist through time. Because of this, we cannot identify traits with the matter out of which they are made; instead, we identify them by the role they play in a functional organization.

Naturalising function ascriptions

Armed with this brief sketch of the organizational account of biological systems, we can now appreciate what it is exactly about biological organization that naturalizes the teleological dimension of function ascriptions. By performing their function, traits that exist in closed systems contribute to the maintenance of a system which, in turn, maintains the trait. Because each trait is mutually dependent on each other, each trait is also dependent on itself. If any trait stops performing its function, it will affect all of the traits that depend on it and, in turn, the traits that depend on them, causing propagating dysfunction throughout the system which, at some point, will loop back around and effect the initial malfunctioning trait. In this way, a trait depends on the effects it creates. A heart only exists so long as it continues to beat. At the core of the organisational account, lies an appeal to the mutual dependency relations which define closed systems.

In addition, the organisational approach circumvents the challenge from backward causation which plagued Wright's account. According to the organisational approach, the effects of a trait really do cause the same token trait to persist through time. Consider trait X which at t_1 causes an effect at t_2 that contributes to maintaining trait X at t_3. Because trait X is the same token trait at t_1 and t_3, we can legitimately say that trait X exists because of the effects it has. If trait X was a different trait at t_1 and t_3, then the effect would be produced by one trait and would contribute to maintaining another trait. If this were true, then the trait and effect would not be *mutually* dependent; it would simply be a linear chain of cause and effect whereby a trait causes an effect which contributes to the maintenance of another trait (Moreno & Mossio 2015, 78). As such, any function ascription which appealed to the effects of trait X at t_2 in order to explain why trait X at t_1

exists, would indeed entail some kind of backward causation. By contrast, the organisational approach instead appeals to some kind of 'quasi-holism' whereby the components contribute to maintaining the whole which in turn contributes to maintaining the components (McLaughlin 2001, 210). This is an issue I discuss in some detail in Article I.

Working within this general framework, different organisational accounts naturalise the teleological dimension of function ascriptions slightly differently. Some hold that it is the past effects of a trait that we refer to when we identify that trait's function (McLaughlin 2001). This gives us a backward-looking organisational account. In this version, a heart has the function of circulating blood because at previous points in time, the heart pumped which caused the circulation of blood, without which the heart would not now be where it is, doing what it doing. Other accounts argue that it is the future effects a trait will have on its own persistence that we refer to when we identify a trait's function (e.g. Schlosser 1998, Christensen & Bickhard 2002), giving us a forward-looking organisational account. In this version, the heart has the function of circulating blood because if it ceased to do so, it would cease to exist. Finally, Moreno & Mossio (2015, 78) have argued that although the relation between a trait and its existence can only be observed in time, we naturalise function ascriptions by abstracting this causal loop from time, giving us an atemporal organisational account. In this case, it is not any specific instance (or set of instances) of the heart beating that we refer to when we say the heart has the function of circulating blood, but an abstract causal regime within which the heart and circulation of blood act as mutually dependent traits (see below).

Now that we have a fuller picture of the organisational approach to functions, it is worthwhile reconsidering the categorization of the organisational approach as an instance of the etiological approach. In many respects, this categorization remains true. Like the etiological account, organisational accounts attempt to account for functions by appealing to a feedback mechanism which explains why the trait exists. But the actual feedback mechanism they appeal to brings the organisational approach well and truly in-line with various versions of the dispositional approach to functions. In particular, it has important parallels to Nagel's account and other cybernetic-inspired accounts.

According to the organisational approach, the closed causal organization of biological systems allows biological systems to maintain themselves through time. It is important to realise that this positions the organization itself as a capacity. Organisational accounts variously identify this as the capacity for self-maintenance, self-reproduction and self-determination (Schlosser 1998, McLaughlin 2001, Christensen & Bickhard 2002, Delancey 2006, Mossio et al. 2009, Mossio & Bich 2017). This makes plain the similarities between organisational and dispositional accounts. Dispositional accounts claim that the functions of traits are defined in terms of the contribution which a trait makes to a particular capacity. The organisational approach claims that functions are defined in terms of the contribution which a trait makes to maintaining the organization. In turn, this organization itself is a capacity, the capacity for self-maintenance, self-

reproduction and self-determination. By tying functions to the contribution a trait makes to maintaining the organization, like dispositional account, organisational accounts tie functions to the contribution a trait makes to a capacity.

In another similarity to Nagel, the organizational approach justifies singling out this particular capacity (the capacity for self-maintenance) as the goal of biological systems on the basis of the system's directed organization. The system will pursue this goal regardless of internal and external disturbances and employs various control mechanisms to do so. But the organisational approach adds an additional condition. As Mossio et al. sum it up: 'A first important implication of organizational closure is that it provides a criterion to determine in a non-arbitrary way what 'the goals' of the system are. The goal states of self-maintaining systems are not such simply because they are characteristic or interesting capacities (as the systemic accounts suggest), or because the systems compensate for deviations from them (as classical cybernetic accounts claim). More fundamentally, in an organizationally closed system the goal states are the stability points (or set of points) through which the system can exist' (2009, 824; see also Barandiaran & Moreno 2008). The point is, we don't attribute goals to just any capacity the system seems organizationally directed towards sustaining. We only attribute goals to systems which are so organised so as to maintain their existence through time.

But this begs a question, why should we interpret systems which are organised to maintain their own existence as goal-directed teleological systems but not systems which are organised so as to achieve some other end? It is this question that I explore in Article I of this dissertation and the answer to it turns out to be intimately connected to the way organisational accounts naturalise normativity. I believe that, at their core, all organizational accounts ascribe some kind of value to existence itself. For example, Moreno & Mossio (2015) claim that existence has an intrinsic *relevance* for the system. This intrinsic relevance, they claim, determines what norms the system should follow: 'the system must behave in a specific way, otherwise it would cease to exist' (Moreno & Mossio 2015, 71). Which is to say, the heart ought to continue circulating blood because circulating blood is necessary for the continued existence of the whole organisation and thus the heart itself. Given existence has some kind of value for the system (existence is 'intrinsically relevant'), the heart ought to do that which secures its continued existence. In short, the norms of the systems are derived from the intrinsic value ('relevance') placed on existence itself. Likewise, McLaughlin (2001) appeals to the intuition that existence is *good* for living systems and that traits that contribute to maintaining the system's existence are therefore *beneficial* for the system. Regardless of the specific terms they use, both approaches are very similar. Because the capacity (self-maintenance) which they contribute to has some value for the system (expressed as either goodness or relevance), this makes traits which contribute to that capacity normative. In this way, the organizational account of biological teleology and normativity is markedly different from Nagel. Where Nagel did not require that the 'characteristic activity' of the system was good for the system, according to the organizational approach, it is the fact that the characteristic activity

is good for the system that makes the characteristic activity a goal of the system. In this way, the intuition that existence has some value underpins both the organisational account of teleology and normativity.

To sum, the organizational approach naturalizes functions by appealing to the mutual dependence that arises between a trait and its effect within biological systems. It argues that all functions must contribute to maintaining the organization on which its existence depends. This organization itself is interpreted as the capacity for self-maintenance; it enables the system to persevere through time. The identification of this end as the end towards which all functional traits must contribute is underpinned by a value placed on existence itself. In turn, the value placed on existence justifies talking about traits which contribute to the maintenance of the system's existence in normative terms. Armed with this interpretation of the organizational approach, I will now consider the significance of two counter-examples which have been raised in objection to the organisational approach. The first type suggests that the organisational approach is too inclusive, it ascribes functions to nonfunctional self-maintaining systems. The second type suggests that the organisational approach is too exclusive, it fails to ascribe functions to reproductive biological traits.

The challenge of simple self-maintaining systems

As discussed above, Wright's etiological account of functions turned out to be too liberal when confronted with simple self-reproducing systems, what are generally known as Boorse-type counterexamples. Recently, Garson (2017) has argued that the organisational account suffers from the same problem.

Consider again the example of the stick which gets stuck against the rock, thereby creating a vortex which keeps the stick stuck against the rock. The crux of the problem is that the 'trait' (defined as the stick stuck against the rock) is dependent on the effect it creates (the vortex) and, in turn, the effect is dependent on the 'trait'. This makes the trait and its effect mutually dependent. If this is true, then any account which appeals to mutual dependence between a trait and its effect to ground function ascriptions seems to be forced to ascribe the stick the *function* of maintaining the vortex. And this is exactly what the organisational account does appeal to in order to ground function ascriptions: mutual dependency relations. Given this, it might seem that the organisational account must ascribe functions in Boorse-type situations.

But the organisational approach has a few extra weapons in its toolbox than Wright did. The organisational approach claims that the trait must exist within a specific type of organization. Mutual dependency is one feature of this organization, but in addition, the system must possess regulative mechanisms, be constituted by functionally differentiated parts and it must create its own boundary systems. None of these criteria seem to be satisfied by the stick and the vortex or other simple self-maintaining systems.

That said, Garson (2017) has argued that the organisational approach is vulnerable to other, more complex, Boorse-type situations. The example he gives is panic disorder, which is a disorder that reproduces itself

through various mechanisms. For example, persons with panic disorder may have increased anxiety that a panic attack will occur and may exhibit certain avoidance behaviours, both of which can trigger further panic attacks, leading to increased anxiety and further avoidance behaviours, leading to further panic attacks, and so on. Garson argues that this disorder seems to involve mutual dependence, functional differentiation and various regulative mechanisms which maintain the disorder. Accordingly, the organisational approach should interpret 'traits' of the panic disorder such as 'hypervigilance to bodily sensations' as functional for the role they play in maintaining the panic disorder.

Regardless of what you make of Garson's presentation of the case (I find it somewhat vague), there is a fundamental problem with the example he provides. The functional trait in question is not 'hypervigilance to bodily sensations'; if it is anything along these lines, it should be rightly categorized as 'vigilance to bodily sensations'. And while it might be true to say that, in the case of panic disorder, the trait is not functioning very well, that is only because in its 'hyper' state, the trait is malfunctioning. But it is not true to say that the trait is nonfunctional. That would be like saying that the heart which beats irregularly is nonfunctional. The point is that malfunctional traits are a sub-class of functional traits. Just as it is true that all traits with a function can malfunction, so too is it true that any trait which is malfunctioning must have a function to malfunction from.

Presumably, Garson does not wish to claim that this trait is nonfunctional and organisational accounts mistakenly attribute functionality to it. Rather, what Garson's example actually points to is the fact that the basic organisational account does not discriminate between traits that function well and traits that function poorly. According to the organisational approach, any trait which contributes to the maintenance of a closed system of which it is a part is functional and their normative status stems from the contribution they make to maintaining the system of which they are a vital part and which, in turn, maintains them. This is the same as the welfare normativity that Hempel appeals to.

What the organisational approach needs is a further set of norms which it can appeal to in order to distinguish between well-functioning traits and malfunctional traits. Currently, it does not distinguish the performance norms which Nagel is concerned with. And indeed, this fact has been explicitly recognized by some proponents of the organisational approach. As Moreno & Mossio put it, the distinction between (well-)functions and malfunctions requires an additional set of norms, on the basis of which it might be possible to discriminate between different ways of contributing to the maintenance of a closed organisation' (2015, 82). In addition, some headway has already been made in formulating these secondary biological norms (see Moreno & Mossio 2015 and Saborido et al. 2016).

The challenge of cross-generation traits

From the very first proposal (Schlosser 1998), proponents of the organizational approach have realized that their accounts suffer from a problem. Namely, it is not clear how the organizational approach can ascribe functions to cross-generation traits such as semen production. Organizational accounts stipulate that functions of traits contribute to the self-maintenance of the system they are part of and, thereby, contribute to their own self-maintenance. But, reproductive traits are produced and maintained by one system in order to produce or maintain another system. As a consequence, the basic organizational approach seems to imply that cross-generation traits, including all reproductive traits, have no function. As Artiga & Martinez sum up the problem, 'in general, organizational accounts fail to attribute functions to any trait whose contribution is realized in individuals other than the bearer of the trait' (2015, 110). Unsatisfied with this result, various organizational accounts have pursued two different methods for dealing with the problem of cross-generation functions.

The splitting account

Although both Schlosser (1998) and Christensen & Bickhard (2002) recognized the problem which cross-generational traits posed for the organizational approach, it was Delancey (2006) who fully dissected the problem and set the framework for future discussion. He argued that the ontological difference between intra-generational traits and cross-generational traits was so great that the organizational approach would have to be re-framed as a disjunctive account[15]

According to this approach, some traits acquire functions in virtue of the way they contribute to the self-maintenance of a system they are part of and thereby contribute to their own self-maintenance. Such traits are intra-generational traits, of which the heart is the classic example. Other traits acquire their functional status by virtue of the way they contribute to the reproduction of their kind. Such traits are cross-generational traits, of which semen production is the classic example. While intra-generational traits reproduce themselves within the lifespan of an organism, cross-generational traits arise when a trait causes its own reproduction in subsequent generations. Significantly, in order for this approach to work, the organizational account must introduce a distinction between traits and tokens. A token cross-generation trait contributes to the maintenance of a system which reproduces another token of the same type.

There are a few points to note about this approach. First, by introducing the type-token distinction, organizational accounts cannot appeal to a causal loop in order to naturalise the teleological dimension of function ascriptions. As Delancey himself acknowledges (2006, 88), while you might metaphorically say that

[15] Schlosser can also be read as endorsing a disjunctive account (1998).

a type reproduces itself, this is and can only ever be a metaphor. When you consider the actual causal relations that exist, one particular reproduces another particular of its own kind. There is no a causal loop and without a causal loop, there is no self-reproduction. Given there is no self-reproduction, we cannot appeal to self-reproduction in order to ground function ascriptions. This means that the type of functionality possessed by intra-generational and cross-generational traits is markedly different. Second and related, by introducing the trait-token distinction, it becomes increasingly difficult to discern the difference between the organizational approach to cross-generation traits and etiological accounts (Artiga & Martinez 2015). Because of these drawbacks, various proponents of the organizational approach have preferred to pursue a unified approach.

The unified account

The alternative approach suggests that one unified organizational account of functions can be applied to both intra-generation and cross-generation traits. This approach was originally suggested by Christensen & Bickhard (2002) and has since been much developed in Saborido et al (2011) and Mossio & Saborido (2016). At its core, the approach makes two ontological claims that allow it to treat cross-generation traits and intra-generation traits the same.

The first claim concerns the system which the reproductive trait contributes to. It claims that the reproducing and reproduced trait form an encompassing system ('a minimal lineage') which is a closed self-maintaining system. As Mossio & Saborido put it, 'our proposal relies on the idea that cross-generation functions are subject to organizational closure just as intra-organismal ones; yet, the system that realizes the relevant closure goes beyond the boundaries of individual organisms' (2016, 266). Importantly, in this system, the parent and offspring are not viewed as two tokens of a type. Instead, this approach views the parent and offspring as two time slices of one token self-maintaining system. To support this claim, Saborido et al. (2011) emphasize the fact that we regularly consider biological systems as the same systems despite the fact that they are constantly changing, materially and structurally. We do this every day when we consider an organism to be the same individual today as it was five years ago. And just as we consider an organism to be the same individual now as it was five years ago, so too should we consider a grandmother, mother, and grandchild to be the same individual. To use their terms, all biological organisms are time-slices of a 'self-maintaining organisation whose extension in time goes beyond the lifespan of individual organisms' (Saborido et al. 2011: 599). The second ontological claim follows directly from the first. If the system extends beyond the boundaries of a single lifespan, so too must the traits out of which that system is constituted. In this way, me, my mother and my grandmother all possess the same heart, the same lungs, the same eyes and, most importantly, the same ovum.

If both these claims are true, then the organizational account can ascribe functions to cross-generation traits in exactly the same way as they ascribe functions to intra-generational traits. Reproductive traits contribute

to maintaining a self-maintaining organization (the minimal lineage) of which it is a vital part and on which its existence depends. If the ovum (extended through the generations) stopped contributing to the maintenance of that system, the organization would cease to exist and so too, would they as traits.

By and large, most other authors have resisted the unified account (e.g. Delancey 2006, Artiga & Martinez 2015, Garson 2017). For many, the encompassing system it appeals to seems both ontologically problematic and theoretically unnecessary. In addition, the claim that the traits which comprise these systems are all instances of the same token trait rather than different tokens of the same type has proved a particularly difficult pill for many to swallow. On both these points, Mossio & Saborido may be right, those offended might simply lack a sufficient understanding of the nature of closure (2016, 270). Though if so, the onus is surely on Mossio & Saborido and their collaborators to clarify the matter.

In any case, it is not my intention to cast judgement on the particular case of reproductive traits here. Rather, my interest lies in a different class of encompassing systems to which the organizational account also appeals. These include bog-standard systems that biologists actually refer to in functional ascriptions all the time: multi-cellular organisms, colonies, and symbioses. As such, in these cases, it is not the appeal to the encompassing system which is problematic for the organizational approach in these cases. And yet, as I argue in Article II and III, the organisational approach still has problems ascribing certain traits in these systems functions and has so for largely the same reason that it struggles with reproductive traits.

Indeed, the particular problem which much of this dissertation drives at has, to some degree, already been foreshadowed by Artiga & Martinez (2015) in their critique of Saborido et al. (2011)'s treatment of cross-generation traits. This somewhat underappreciated criticism goes as follows.

'We may now note that those traits whose function, organizational account wants to claim, depends on their role in the maintenance of a single individual – e.g., hearts, kidneys, brains, etc. – also participate in the maintenance of the same lineage-encompassing system: it is obvious that donkey hearts have contributed to the maintenance of the donkey lineage: without hearts, the individuals that form part of the lineage would not have existed. It is equally obvious that the donkey lineage has maintained and produced hearts: without offspring donkeys there would have been no new hearts, just as there would have been no new semen' (Artiga & Martinez 2015, 115)

There are a number of problems which this treatment of the unified account. Most pointedly, the emphasis on new hearts, new semen and individuals seems rather to miss the point of the proposal entirely. But, despite this, I believe that there is an important truth to the criticism raised. If a trait simultaneously exists within two closed systems, how do we know which closed system grounds its functional status? On what grounds is the function of semen grounded in the encompassing system while the function of the heart is grounded in the

individual organism? Bearing this general question in mind, I will now introduce the articles that form the core of this dissertation.

Contributions of this dissertation

This dissertation arose out of an interest in the machine analogy of the organism. Specifically, I have sought to understand exactly what it is that lies at the heart of this analogy and how strong that foundation is. I have presented good reasons why I believe that it is the functional status of both machines and organisms that drives the analogy. And I have argued that explicating the nature of biological functions will give us some basis from which to cast judgement on the analogy itself. If it is true, as many claim, that the functional nature of organisms and machines is ontologically different, then the analogy is at best, an analogy. While there might be some very important similarities between the two types of systems, there are also some very important differences. They are not one of a kind.

After considering a number of different accounts of functions that have been proposed over the last sixty years, all of which have strengths and weaknesses, I have elected to focus on the organisational approach to function ascriptions in my research. This is not to deny that the etiological and dispositional accounts have something to say about the machine analogy, they do, but I believe there are good reasons to concentrate on the organisational approach instead. First, it is new and ambitious and, as such, there is much worth saying about it. Second, I believe it taps into an important truth about the way biologists actually ascribe functions. Third, proponents of the organisational account (and philosophers who have been influenced by them) are wont to make a number of claims about the machine and organism analogy. McLaughlin has claimed that only organisms 'have a good', meaning that only organisms can be beneficiaries from function ascriptions and that this fundamentally distinguishes them from machines. Nearly all proponents of the organisational approach have claimed that biological function ascriptions are uniquely biological and are subject to different truth conditions than artifactual function ascriptions. Furthermore, following from this, a number of authors (McLaughlin 2001, Nicholson 2013, Mossio & Bich 2017) have argued that the purposive nature of organisms and machines is fundamentally different; where machines are extrinsically purposive, organisms are intrinsically purposive. Each of the articles in this dissertation takes aim at a different one of these claims. But before introducing each in turn, it is worth saying something about the methodological approach that I adopt in these articles.

Methodological considerations

In this introduction, I have considered a number of different accounts of biological functions underpinned by a variety of different methodological beliefs. Hempel and Nagel's account were developed within the Deductive-Nomological tradition, an approach which all later accounts reject. And while most later accounts are either explicitly or implicitly engaged in a project of conceptual analysis, two of the most important accounts – Millikan's and Cummins' – both explicitly reject the conceptual analytical approach. Given mine

is indeed a project of conceptual analysis, before getting started, it is worth saying something about why I have chosen this methodology and the type of conceptual analysis with which I am engaged.

At its most basic, conceptual analysis is concerned with trying to understand what something is (Glock 2017). When a scientist utters the word 'function', it is the task of the conceptual analyst to identify what they mean. What beliefs, ideas, perceptions, biases, have led them to utter this particular word in this particular circumstance? (Neander 1991). The output of this project is usually the identification of truth conditions for the application of the term. Although it is undoubtedly still 'the most widely used tool in the analytic philosopher's toolkit' (D'Oro & Overgaard 2019, 6) and is certainly the dominant approach in the analysis of functions (Allen & Neal 2019, 7-8), in recent years, the entire method of conceptual analysis has come under increasing scrutiny.

As Stitch (1996, 171-174) describes it, what actually happens when philosophers engage with conceptual analysis goes something like this. First, one philosopher identifies a set of necessary and sufficient conditions for the application of a term like 'function'. Then, some other philosopher comes up with some counter-example that disproves those truth conditions and, instead, proposes a new set of truth conditions. These are, in turn, knocked down by some new counter-example that some other philosopher comes up with, and so on and so on. There are two features of this practice which Stitch has argued are particularly problematic. The first problem with the approach is its reliance on necessary and sufficient truth conditions which, most agree, are nearly impossible to formulate. The second problem with this approach is its reliance on intuitions, which often conflict. As Bigelow & Pargetter famously characterize the problem, 'The test of examples and counterexamples is important. Yet in this case, in the analysis of functions, there is a risk that it will decay into the dull thud of conflicting intuitions' (1987, 196).

In Article II and Article III of this dissertation, I discuss various types of biological interactions and argue that the organisational approach cannot ascribe them functional status. In parallel, I argue that any account of functions should be able to ascribe them functional status. As such, I present these biological interactions as possible counter-examples (or at the least, problem areas) for organisational accounts. Does this mean that I am engaging in the kind of 'armchair philosophy' so derided by Stitch and which, for good reasons, has generally fallen out of fashion?

As I see it, the short answer to this question is no. Regarding the first complaint, it is not necessary to my project that the conditions for function ascriptions which I discuss should be interpreted as necessary and sufficient truth conditions. For example, I could just as easily adopt the approach proposed by Neander to instead seek out 'criteria of application that people standardly apply, most of the time, in most standard contexts' (1991, 171). Neander's intention here is to limit the relevance of near-impossible fictitious counter-examples such as Davidson's swampman and I think there is much to be said for Neander's reasoning in this

case. But regardless, it does not make much difference to the counter-examples which I present. Not only are the counter-examples that I present very real, they are also very common. Indeed, I would go so far as to say they are a fundamental feature of biological organization and, as such, any account of biological functions based on organization should be able to accommodate them.

Nor do I think the second complaint has much bearing on my methodological approach. Quite simply, I do not appeal to intuitions, I appeal to empirical case studies and I make a careful study about the way biologists actually describe those case studies. Very specifically, I make a study of the way biologists apply the term 'function' within these contexts and what they mean when they do. In this way, my methodological approach is naturalistic, it embraces the idea that philosophy must engage with empirical science (Kornblith 2019). When I analyze the phenomenon of apoptosis and argue that the organisational account of function does not apply to it, I engage with what Glock calls 'impure conceptual analysis' (Glock 2017). I blur the boundaries between conceptual analysis, which involves identifying conditions of application, and empirical science, which 'determines whether or not these concepts do apply to organisms or systems of a certain type; (and) also provides causal explanations of how the instances of these concepts come to satisfy these conditions of application' (Glock 2017, 80). I am not suggesting that we blindly take what biologists do and say as gospel, I am not suggesting biologists can't be wrong. But I do believe that it is very difficult to do good conceptual analysis without engaging with empirical science. Nowhere is this truer than in the philosophy of science.

Having given my response to the standard negative arguments typically raised against conceptual analysis, I now turn to a few positive arguments as to why I believe conceptual analysis is the right methodological approach in this particular circumstance.

First and foremost, I am actually interested in what biologists mean when they use a term like functions. I am also interested in what philosophers mean and what laymen mean. I think all these perspective bring valuable insight into the nature of things, as too can the differences that arise between them.

Second, I believe that a consideration of counter-examples has actually made a valuable contribution to our current understanding of what biological functions are. As the historical survey presented above attests to, truth conditions can be improved. I think there is no doubt that Wright's account was an improvement on Hempel's. Nor do I think there is there any doubt that Millikan and Neander's accounts are an improvement on Wright's. Likewise, I think both of these accounts are improved by Godfrey-Smith's contribution. Although no etiological approach has proved perfect, over-time, they have become less problematic, more comprehensible, and the number of counter-examples they have faced has been greatly reduced. Much of this progress, I would argue, has arisen through a consideration of counter-examples that afflicted previous accounts.

Third and finally, all proponents of organisational accounts are either explicitly or implicitly engaged in conceptual analysis. They don't shy away from nominating various truth conditions, nor do they shy away from evaluating those truth conditions by considering various counter-examples and problem cases. This is what the entire debate about cross-generation traits is. Ultimately, it is my hope that some of the research in this dissertation might enrich the organisational approach, that it might lead proponents of organisational accounts to either reconsider the truth conditions they appeal to or the way they apply them. For my research to achieve this goal, it will be helpful to share the basic assumptions and methodological approach shared by other proponents of the organisational approach.

Article overview & final reflections

Article I: The Goodness of Preservation in McLaughlin's Account of Functions

In this article, I analyze a core claim of McLaughlin's organisational account of functions: that biological systems are beneficiaries and it is this fact that grounds their functional status. Ultimately, I argue that McLaughlin's account turns on the intuition that preservation is intrinsically good. It is this assumption, rather than facts about biological systems, which grounds both the teleology and normativity in his account. This raises questions about the status of McLaughlin's account as a naturalistic account of functions. In addition, appealing to this intuition directly conflicts with two further claims McLaughlin goes on to make. First, it undermines his claim that whole organisms don't have a function and, second, it undermines his ambition to develop an account of functions which appeals to formal causes in place of final causes.

Because McLaughlin assumes that preservation is the only good for biological systems, he also becomes committed to the notion that this is their only goal. This has consequences for the way McLaughlin characterizes functional traits: McLaughlin claims that only traits which contribute to their own self-reproduction are functional (self-reproduction here should be interpreted as self-maintenance via self-regeneration). But this raises questions about the functional status of a variety of biological traits that, quite blatantly, seem to function for ends other than their own self-reproduction. As I discussed above, a similar objection has previously been raised against the organisational approach concerning reproductive traits. This article contributes to that debate by identifying the underlying root cause of this bias: organisational accounts fail to ascribe reproductive traits a function because proponents of organisational accounts have assumed that self-preservation is the only good and only goal of biological systems. In the following articles, I make further contributions to this debate by showing that reproductive traits are not the only traits which the organisational approach struggles with. In addition to reproductive traits, organisational accounts struggle to ascribe functional status to a variety of competitive and cooperative relations that arise between biological systems.

Before introducing those articles, it is worth noting that although I take McLaughlin's account as my target in Article I, the arguments I present can be directed against other versions of the organisational approach as well. As I discussed above, the organisational account put forward by Mossio et al. (2009), Saborido et al. (2011), Moreno & Mossio (2015) and Mossio & Bich (2017) appeals to a similar value judgement about preservation. In their account, that value is expressed as intrinsic relevance rather than intrinsic goodness but the basic idea is the same; preservation has some kind of value and it is this value judgement which justifies talking about functional traits in normative terms. In addition, I have argued above, Mossio et al. (2009) appeal to this value in order to single out self-maintenance as a *goal* of the system. While other ends a system serves are interpreted as capacities of the system, only the capacity of self-maintenance is interpreted as a goal. I have argued that the only justification for this move lies in an appeal to the intuition that there is something special about preservation that singles it out and places a value on it which we would not place on other ends. While such a claim might be justifiable in ethics, there is nothing in the organization of biological systems itself that justifies this move. As it is in McLaughlin's account, so too is it here: Mossio et al. (2009) appeal to the intuition that preservation if good in order to ground both the teleological and normative dimensions of function ascriptions.

Finally, it is worth reflecting on the implications which the arguments in this paper have for the machine analogy of the organism. As I note in the article itself, McLaughlin's approach to naturalizing teleology and normativity seems to imply that various physical dissipative systems are also normative and teleological systems. This is a point which is explicitly endorsed in Bickhard (2000, 114), Mossio et al. (2009) and Moreno & Mossio (2015, 71). Because normativity and teleology is grounded in self-maintenance, and because self-maintaining systems exist in both the physical and biological world, then it follows that physical and biological systems share the same kind of normative and teleological status. Mossio & Bich have defended this consequence, claiming, 'This result seems at odds with both scientific practice and intuition. In reply, we answer that, indeed, if self-constraint were shown to exist beyond the biological domain, the organisational account we advocate should concede that teleology exists beyond biology' (2017, 1091).

But although they concede that normativity and teleology exist beyond the biological domain, they deny that purely physical systems are functional, arguing that functionality requires an extra dimension over and above normativity and teleology; it also requires a certain level of complexity. They argue that physical dissipative systems are single-monolithic systems which lack differentiated parts that functions can be ascribed to. This makes some headway into restricting functional status to the biological domain. But, it is important to realise that this is not an argument against the possibility of non-biological systems possessing the same functional status as biological systems. Rather, it is simply a claim that there are currently no known examples of such systems. This opens the door to the possibility that in the future, such systems will either be discovered in the natural universe or might be created by man. Indeed, this is a point which McLaughlin himself explicitly acknowledges. As he puts it, 'the fact that the machines that we have thus far produced only have external

functions and are thus only good for us (or whoever we have in mind) but have no good of their own is a purely contingent state of affairs' (2001, 185).

This suggests that, there is no reason in principle why one day we could not build machines which are beneficiaries, with a good of their own and possessing the same normative, teleological and functional status as organisms. If this is true, then it seems the prospects for the machine analogy of the organism are bright and that one day, soon perhaps, the machine analogy of the organism will recover the position endowed on it by Descartes. It suggests that once again, the machine analogy might come to be seen as an ontological claim about the nature of organisms and machines rather than just an extremely useful analogy and heuristic tool.

Article II: The Challenge of Apoptosis for the Organisational Account of Functions

In this article, I turn to a set of questions which were raised but not answered in the previous article. Is it true that all biological traits function for their self-preservation? Is it true that persistence is the only good and goal of biological systems? I explore this question by examining whether the organisational account of functions proposed by Mossio et al. (2009), Saborido et al. (2011), Moreno & Mossio (2015), Mossio & Saborido (2016) and Mossio & Bich (2017) is equipped to ascribe apoptotic traits a function.

The organisational account of functions claims that functional traits are those traits which contribute to the maintenance of a system which, in turn, maintains the trait. Apoptosis is the process by which cells actively participate in their own annihilation, for the benefit of an encompassing system they are embedded within. Apoptosis presents an obvious problem case for the organisational account, self-annihilation is quite clearly a different end from self-maintenance. It also points to a wider problem which the organisational account has with regard to altruism; if everything functions for self-maintenance, the possibility of traits that function in order to other-maintain does not arise. Given these problems, I suggest that the organisational account should adopt Delancey's (2006) proposal and re-frame itself as a disjunctive account in which traits can either function to maintain a system they are part of or to maintain another system of the same type. I argue that the disjunctive approach is necessary as it makes room for the fact that biological systems must behave differently when they are simultaneously an autonomous system and a part of an encompassing autonomous system.

One consequence of the analysis offered is that it requires a new interpretation of the nature of biological teleology. The organisational account offers an interpretation of the teleological nature of biological systems which grounds that teleology in self-maintenance. But my analysis – and Delancey's account which I appeal to – explicitly rejects the claim that all biological traits function for their own self-maintenance. This raises an obvious question: if teleology is not grounded in self-maintenance, what is it grounded in? What kind of teleology is it? What relation does it have to other types of teleology? These are the questions which I take up in the third and final article of this dissertation.

Article III: Extrinsic Teleology in Symbiotic Organisation

In this article, I analyse the claim oft-made by proponents of the organisational approach to functions that biological teleology is intrinsic (e.g. McLaughlin 2001, Nicholson 2013, Mossio & Bich 2017). I start by identifying the features of a biological trait which make it intrinsically teleological, within the organizational approach. Put roughly, intrinsically teleological traits are those traits which contribute to maintaining a system of which they are a part and on which their existence depends. I then turn to the question of whether it is true that all biological traits are intrinsically teleological. After considering a range of symbiotic interactions and discussing the types of teleological relations that arise within them, I argue that the functional traits of symbionts can be both intrinsically and extrinsically teleological. Furthermore, extrinsically teleological traits can arise through either competitive or cooperative interactions. These two types of interactions lead to two very different types of extrinsic teleology. While cooperative extrinsic teleology does not seem particularly problematic for the organizational approach, it seems impossible for the organisational approach to accommodate competitive teleological relations.

The analysis offered in this article brings together a number of different threads which run throughout this dissertation. First and foremost, it immediately picks up the challenge which was raised in Article II. There (as here) I have argued for the existence of functional traits which serve ends other their own self-maintenance. This raised a question about the type of teleology that these traits possess. I have responded to that challenge with an account of extrinsic teleology rooted in the competitive and cooperative nature of biological systems.

There are many questions which remain. The most important of these, I believe, concerns the relation between competitive and cooperative extrinsic teleology. I suggest that one useful model for thinking about this relation can be found in the account of biological collaboration put forward by Dupre & O'Malley (2009). By this account, living systems are not just distinguished by the fact that they are self-maintaining, but also by the fact that they are collaborative. Collaboration here includes behaviours that are both competitive and cooperative (as well as mutualistic and commensal). This suggests that all extrinsically teleological relations are collaborative. Note that this is not a suggestion that all collaborative relations are teleological - many are not - but some are. Roughly put, the gist of the idea is that intrinsic teleological relations function to sustain self-maintaining systems via autonomous organisation and extrinsic teleological relations function to sustain self-maintaining systems via collaborative organization.

Finally, it is quite obvious that the arguments presented in this article will have consequences for the prospects of the machine analogy of the organism. At first glance, it may appear that the analysis suggests that the functional nature of machines and organisms is more similar than proponents of the organisational account have claimed. They have claimed that the teleology of organisms is intrinsic while machines are

extrinsically teleological, as such, the purposiveness of machines and organisms is fundamentally different. By contrast, I claim that the teleology of organisms can be both extrinsic and intrinsic. This seems to imply that organisms and machines have more in common than previously proposed.

But considered through the lens of collaboration, the consequences of my analysis become somewhat more complex. As noted above, the organisational approach implies that if we can build a self-maintaining machine, organisms and machines will have the same functional nature. But I have claimed that organisms do not just have an intrinsic telos – they have this – but in addition, they also have a capacity to collaborate, to re-purpose themselves and other biological systems around them. This makes the gap between machines and organisms somewhat wider than proponents of the organisational account have claimed. In order for the machine analogy to reclaim its ontological standing, not only would we have to build a self-maintaining system but that self-maintaining system would also have to be able to interact with other self-maintaining systems and, for better or worse, effect their self-maintaining organization.

ARTICLE I: THE GOODNESS OF PRESERVATION IN McLAUGHLIN'S ACCOUNT OF FUNCTIONS

Introduction

This paper examines the organisational account of biological functions proposed by McLaughlin (2001). The gist of the account claims that functions of traits contribute to the self-reproduction of a system that the trait is a part of and on which its existence depends. In this way, McLaughlin's account is one example of the organisational account of functions (Schlosser 1998, Collier 2000, McLaughlin 2001, Christensen & Bickhard 2002, Delancey 2006, Edin 2008, Mossio et al. 2009, Saborido et al. 2011, Moreno & Mossio 2015, Mossio & Saborido 2016, Mossio & Bich 2017). But the account proffered by McLaughlin is unique. Unlike other organisational accounts, it characterizes functions as good and the biological systems which have them as beneficiaries.

Immediately, this raises a number of important questions about what it means to be good and what it means to be a beneficiary. In this paper, I argue that the answers McLaughlin gives to these questions are problematic. At its core, his account must appeal to the assumption that existence is intrinsically good and this assumption is not grounded in any objective features of biological systems but is grounded in the beliefs of biologists ascribing functions. In turn, this seems to undermine any claim that McLaughlin's account has provided a naturalized account of biological functions. [16]

Even putting aside the issue of naturalism, this approach raises a number of questions about the internal consistency of McLaughlin's account. McLaughlin's implicit appeal to the goodness of preservation directly contradicts a number of explicit claims he makes elsewhere. First, it re-introduces final causes into biological explanations and second, it suggests that whole organisms (not just their parts) have functions. In addition, it raises some questions about how the account would deal with a range of biological phenomena including altruism, symbiosis and nested hierarchies. Finally, as I interpret him, McLaughlin introduces the appeal to

[16] To be fair, McLaughlin never explicitly claims to have naturalised normativity. That said, his project is thoroughly naturalistic in spirit. Starting from a consideration of functional explanations, his objective is to make plain the metaphysical commitments which such explanations entail. As to the question of whether we should accept those consequences and, therefore, the explanatory practices themselves, McLaughlin mostly remains agnostic. That said, he does claim that 'there appears to be nothing radically incompatible with naturalism in the commitments demanded by functional explanation' (2001: 212), which seems to suggest that, to some extent at least, he perceives himself to have offered a satisfactory naturalization of biological functions.

beneficiaries in order to deal with a number of otherwise pesky counter-examples. That is, he introduces to preserve a proper distinction between functional and nonfunctional systems. Rather than achieve this, the account of beneficiaries McLaughlin actually appeals to ends up ascribing beneficiary and functional status to a range of physical dissipative systems.

Finally, it is worth noting that although I do not argue for it in this paper, it seems likely that the present analysis would have some bearing on how we understand other organisational accounts. Although it is true that other organisational accounts don't appeal explicitly to benefit and beneficiaries, all organisational accounts are normative and it is widely accepted among proponents of the organisational account that they all adopt something like McLaughlin's reasoning in order to justify a naturalized grounding for biological normativity (Saborido et al. 2011, 589).

This paper is spilt into four sections. Following the introduction, in the second section, I provide the necessary context and background, explicating McLaughlin's theory of biological organisation. In the third section, I outline McLaughlin's account of functions. In the fourth section, I identify McLaughlin's motivations for appealing to benefit and beneficiaries in the first place. In the fifth section, I explicate McLaughlin's account of beneficiaries. In the sixth section, I argue that the appeal to beneficiaries conflicts with several other claims that McLaughlin elsewhere makes. Finally, in the seventh section, I offer a summary and conclusion of the arguments in this paper.

Mclaughlin's account of a self-reproducing system

McLaughlin's account of beneficiaries is rooted in his account of biological organisation. As such, in order to say something about his account of beneficiaries, it is first necessary to say something about his account of biological systems.

McLaughlin argues that biological systems can largely be defined by two key features. The first concerns the identity conditions of biological systems and the second, the unique causal structure of those systems. As McLaughlin puts it plainly, to accept functional explanations, 'we must be prepared to accept some rather unusual entities with somewhat peculiar identity conditions, not to mention some somewhat unorthodox internal causal relations' (2001: 173). Both of these two ideas come together in his theory of the organism as a self-reproducing system.

The identity conditions of biological systems

It is a core tenant of McLaughlin's account that biological systems have unique identity conditions which distinguish them from non-living systems. To comprehend the point clearly, it is worth considering what is meant by the term self-reproducing. What is *not* meant is a process of reproduction in which a new individual is made, such as happens when two organisms sexually reproduce. What *is* meant is the process

by which all biological systems constantly reproduce themselves by replacing the cells out of which they are made. In this way, I am constituted by very different matter today than I was, say, ten years ago. I am also structurally very different from the system I was ten years ago. In these senses, it seems fair to say that I am a different system to the system I was ten years ago. And yet, common sense tells me that there is a sense in which I am the same system. In some fundamental way, despite the fact that most of the matter out of which I am made might change, as well as my structure and morphology, I remain the same. As McLaughlin puts it, 'A system is self-reproducing in this technical sense only if its identity over time is constituted by the activity of constantly replacing its parts. It stays the same by rebuilding or reproducing itself' (2001: 211). The meaning, then, of self-reproduction sits closer to the modern idea of self-regeneration and self-repair.[17]

But the force of the insight it is not merely that I remain the same *despite* these changes: in fact, I remain the same *because* of these changes: an organism 'is a system that remains identical to itself *by* renewing its parts and assimilating anorganic matter to its organic structure' (McLaughlin 2001: 173). To put it bluntly, if cell division ceased and the constituents of my body stopped being replaced, many of my vital processes would quickly stop working and I would die. In this way, the identity conditions of all biological systems are dependent on processes through which their material constituents are constantly repaired and replaced. *In order* to remain alive, their material parts *must* constantly change. [18]

The unique causal structure of biological systems

It has long been argued that biological systems have a unique causal structure which, at the very least, *seems* to be different from the causal structure of other physical systems. McLaughlin characterizes this causal structure as self-reproducing: 'The organism reproduces itself by reproducing its parts, and the parts reproduce themselves by contributing to the reproduction of the organism' (2001: 172-3). The emphasis, then, is on a causal loop in which the whole is causally dependent on the parts and the parts are causally dependent on the whole. One half of this claim is relatively uncontroversial; the idea that the existence of the whole is dependent on its parts is entirely in keeping with the dominant mechanistic approach to science. But the other half of this claim, the idea that the existence of the parts is dependent on the whole, is more problematic. One of the problems is that it seems to imply backward causation, because the system, as a whole, does not exist prior to its parts. McLaughlin argues that we can make sense of this if we consider the fact that the whole is constantly regenerating its parts:

[17] For a full discussion, see McLaughlin 2001: 173-179

[18] E.g.: 'such systems remain identical to themselves over time only be re-producing their parts. Token systems that loose the ability to do this don't just break up, they die' (McLaughlin 2001: 182); see also McLaughlin 2001: 13.

> 'A system that remains the same only insofar as it re-produces itself by renewing and replacing its own parts is also temporally prior to (many of) those parts and can thus, without backward causality, be held to be causally responsible for the existence and/or properties of those parts' (2001: 173).

Put bluntly, we can evade the charge of backward causation by appealing to the unique identity conditions of biological systems. On the one hand, the whole system, its sub-systems and organs remain the same. On the other, at some level of description, we do talk about the creation of new biological parts, for example, we talk about new cells emerging out of the process of cell division. As such, it seems fair to say that the whole organism, its nervous system, or its heart exists prior to its youngest cells. As such, we can appeal to the heart to explain properties of the cell, without entailing any backward causation, because the heart exists prior to the cell.

The real problem, then is not backward causation, but downward causation. Claiming that the parts are (causally) dependent on the whole system of which they are a part seems to entail some kind of holism, the ability of the whole to have a real causal influence over those parts. According to McLaughlin, this consequence might be something that we can explain away in the future but it might not.[19] In the end, everything that we know about the way biological systems work might simply force us to accept that holism is true; biological parts have system-dependent properties which we cannot explain without referring to the system of which they are a part. While acknowledging that holism isn't without controversy, largely because of the challenge it presents to mechanism, McLaughlin argues that the truth of holism is ultimately an empirical question, not an analytical one. As such, it is not one that you can oppose *in principle.*[20]

Mclaughlin's account of biological functions

The gist of McLaughlin's account of functions requires that functional traits contribute to the maintenance of the self-reproducing system of which they are a part and on which their existence depends. Specifically, McLaughlin identifies three conditions which must be satisfied in order to ascribe a trait a function:

> 'When we say, with (causal) explanatory intent, that the function of *X* (for *S*) is *Y*, we mean at least:
>
> *X* does/enables *Y* (in or for some *S)*

[19] E.g. 'the implied holism might turn out to be only apparent – though at the moment I see no way to guarantee this' (2001: 173).

[20] E.g. 'While *mechanism*, or causal *reductionism* may in fact, as Kant insisted, be the only game in town with a future, this is nonetheless an empirical not an analytical question' (McLaughlin 2001: 28).

> *Y* is good for some *S*; and
> By being good for some *S*, *Y* contributes to the (re)production of *X* (there is a feedback mechanism involving *Y*'s benefiting *S* that (re-)produces *X*) (2001: 140).

McLaughlin then argues that these three conditions entail three metaphysical commitments. The first condition assumes the existence of causal connections or dispositions. The second condition assumes the existence of beneficiaries, the type of systems which things can be good or bad for. The third condition assumes the existence of some kind of feedback mechanism. Let us consider each condition in turn.

The causal condition *(X* does/enables *Y* (in or for some S*))*

When we ascribe a function to an entity, we assume that it really does cause a certain effect. This is the first condition of McLaughlin's account of function ascriptions. In turn, this condition entails that cause-effect relations really do exist. This is the first metaphysical commitment of function ascriptions. As such, the first condition seems to commit us to nothing over and above the commitments of explanations in the pure physical sciences. That said, the causal relations of biology are of a very peculiar type. Most obviously, functions are non-accidental.[21] Think of an astronaut who accidentally sends a satellite crashing into the Pacific Ocean after throwing his hammer away in a fit of despair. It is certainly not the *function* of hammers to knock down satellites, and yet, it makes perfect sense to say that the hammer *caused* the satellite to come crashing down. The common way of expressing this difference is to say that functions are not merely causes of effects, but mean to ends. Hammers are means to pushing nails into hard surfaces, not for sending satellites crashing in to the earth. What then, makes something a means to an end, instead of a mere cause of an effect?

The first obvious idea is that function bearers cause effects with some kind of regularity: hammers regularly push nails into hard surfaces and have only once (and then, only hypothetically I hope) sent a satellite careening through the stratosphere. But this will not do. There are many causal relations in nature which are regular: every month the new moon rises, every day the earth circles the sun, every evening the tides change; and yet, we hesitate to say that the function of the earth is to circle the sun, the moon to rise, or the tides to change. These seem to be things that the earth, moon and tides just *do*. By contrast, pumping blood is not *just* something that my heart does. It is something that my heart does but in addition, it is something that my heart has the *function* of doing. What then, is the difference between the earth, the moon and the tides on the one hand and my heart on the other? To explicate this difference, the difference between cause and effect

[21] E.g. 'Only if we think we can separate it from accident do we ascribe a function to a trait' (2001: 169); 'the relevant consideration is whether the candidate for the status of function occurs often and regularly enough that its benefit does not seem to be merely accidental. If the effect or its beneficial character seems accidental, we do not ascribe it a function' (2001: 168).

versus means to end, McLaughlin emphasizes two further dimensions of function ascriptions. To sum it up, whenever we ask if a trait has a function, this really boils down to two questions: 'is the trait good for him, and is it there because it is good for him?' (2001: 172). That is, it really boils down to the beneficiary condition and the feedback condition.

The beneficiary condition (*Y* is good for some *S*)

According to McLaughlin, functions are not mere cause of effects, they are causes of effects that are good for some system.[22] But again, this is not enough; it seems perfectly reasonable to argue that the earth circling the sun has a good effect. It is good for the plants that depend on it, it is good for animals like me that depend on the plants, and so on. And yet, it still does not seem right to say that the function of the earth is to circle the sun, for that would seem to imply that the earth circles the sun *in order* to provide energy for life on earth. But as the other planets in the solar system seem to testify, planets will circle the sun regardless of whether it benefits the life that lives on them. Circling the sun is just something that planets do. The problem is that, in order to explain *why* the Earth circles the sun, we need not refer to the benefit provided by that activity to forms of life on Earth. In this way, the second and third conditions are intricately connected. In order to qualify as a function, it is not enough that an effect be beneficial, the beneficial nature of an effect also has to figure in the feedback condition.

The feedback condition (By being good for some *S*, *Y* contributes to the (re)production of *X*)

This brings us to the third condition of biological functions: the teleological idea that functional effects explain a function bearer's existence. Put simply, it seems to be an assumption by both biologists and lay people alike that we have hearts *because* they pump blood and because pumping blood is good for me. By contrast, we do not believe the Earth exists because circling the sun is good for plants and animals.[23] We rightly believe that the earth would circle the sun regardless of whether or not there existed any plants and animals to benefit. As McLaughlin puts it:

> 'Things with functions are thus conceptualized not just as causes of effects but also as means to certain ends. Viewing something as a cause makes no particular presuppositions about why it is there. Although causes are not necessarily there for

[22] E.g. 'Just because something actually works – in other words, has effects – doesn't mean it has a function. Function bearers are means to ends, not just causes of effects: They are cause of valuable or beneficial effects – whatever the source of the valuation' (McLaughlin, 2001: 78)

[23] The common way of saying this is to say that the earth does not circle the sun *in order* to provide energy for the plants and animals, or *for the sake* of providing energy for the plants and animals.

> the sake of their effects, means are in fact there for the sake of their effects' (2009: 96).

To make sense of the teleological aspect of function ascriptions, McLaughlin argues that we need to introduce the notion of a feedback mechanism. When we say that I have a heart because it pumps blood, what we mean is that I have a heart because at earlier points in time that *very same* heart pumped blood, which is necessary for the self-reproduction of the metabolic system of which my heart is a part. This metabolic system, in turn, reproduces the heart which reproduces it; it forms a feedback mechanism. As such, the heart's pumping blood is not just beneficial to the system of which it is a part, it is also beneficial to the heart itself. The beneficial nature of the effect which the heart produces explains why the heart exists. This entire organization is very different from the type of relations which exist between the earth, the sun and the solar system. There is no feedback mechanism by which the solar system reproduces the earth because circling the sun is good for plants and animals. Thus, the earth's circling the sun does not explain why the earth exists and circles the sun.

We are now in a position to appreciate why the earth does not have the function of circling the sun. The earth circling the sun may be beneficial to the life that inhabits it, but in addition to being beneficial, McLaughlin argues that a function must occur *because* it is beneficial. A reasonable question to ask at this point is why appeal to the beneficial nature of functions in the first place. To distinguish the earth which circles the sun from the heart which circulates blood, is it not enough that we simply refer to the feedback mechanism and the teleological dimension of function bearers? Is the normative aspect really necessary at all? With that in mind, let us take a closer look at the motivations for appealing to benefit and beneficiary status.

Why appeal to beneficiaries?

As I interpret him, the appeal to benefit and beneficiaries in McLaughlin's account is necessary to distinguish functional relations from other types of non-functional self-reproducing systems. In short, without the beneficiary condition, the account of functions grossly underspecifies what a functional system is and would have us ascribe functional status to all sorts of unlikely systems, including tumors, flames and hurricanes[24].

In many respects, the organizational account is not completely new. Rather, it offers a revision of the pre-existing etiological account. In particular, McLaughlin's account is heavily influenced by Wright's (1973) and Ruse's (1982) etiological accounts. Such accounts have always been plagued by pesky counter-

[24] See Garson (2017) for a similar interpretation of the role of the beneficiary condition in McLaughlin's account.

examples, most notoriously by swamp-men and simple self-maintaining system. Against all intuitions, etiological accounts seem to imply that the former don't have functions and the later do. It thus seems arguable, that if organizational accounts are to be a genuine step forward for etiological accounts, they must eliminate at least some of the pesky counter-examples which have plagued these accounts. In this spirit, the appeal to benefit can be interpreted as an attempt to delimit the proper class of functions. In particular, it is intended to restrict the ascription of functions to simple self-maintaining systems. To see the problem clearly, consider the following example, originally taken from Van Gulick, which Bedau references:

> 'There is much to be said for this approach to teleology, but as it stands it is not sufficient – something that is not teleological might nevertheless have an etiology like the heart's. Consider a stick floating down stream, which brushes against a rock and comes to be pinned there by the backwash it creates. The stick is creating the backwash because of a number of factors, including the flow of the water, the shape and mass of the stick, etc. But part of the explanation of why it creates the backwash is that the stick is pinned in a certain way on the rock by the water… the backwash keeps it pinned there and being pinned their causes the backwash (Bedau 1992, 648).

The point of Bedau's example is that the stick, like the heart, is where it is doing what it is doing because of the effect it produces. But, it seems a stretch to say that the function of the stick is to create backwash. This seems to suggest that the causal condition and the feedback condition are not enough, on their own, to secure a function ascription. Something must be missing. According to both Bedau and McLaughlin, the evaluative element is missing. Functions do not merely bring about effects which explain the existence of function-bearers, they bring about *good* effects which explain the existence of function-bearers.

Now, the stick in the stream does not seem to present a direct challenge to McLaughlin's account. Clearly, the stick is not *reproduced* by the system comprised out of the stick/rock/backwash. As such, the system is not a self-reproducing system. There are, however, more problematic counter-examples which McLaughlin might have to face up to. In particular, McLaughlin would have to face up to the challenge presented by dissipative systems such as flames. Unlike sticks in streams, flames do seem like potential candidates for being self-reproducing systems: they continuously pull in oxygen and converts anorganic material into fuel, thereby replenishing the material constituents out of which they are made (this example is borrowed from Moreno & Mossio 2015, 71). As such, it at least seems arguable that flames satisfy both the causal and feedback conditions. Without the beneficiary condition, McLaughlin would be forced to ascribe functional status to flames and other self-reproducing dissipative systems.

Unfortunately, even with the beneficiary condition, McLaughlin is going to struggle with counter-examples such as flames. As I will discuss shortly, McLaughlin ties the beneficiary status of systems to their self-reproducing nature. As such, if the system satisfies the feedback condition, the system is a beneficiary and its activity can be described as beneficial, thus the system will satisfy all three of McLaughlin's conditions and seems to be functional under his account. While McLaughlin's account might be able to demarcate the solar system from a cardiovascular system, ascribing functional status only to the later, it cannot demarcate between biological systems and dissipative systems, ascribing functional status to both.

McLaughlin's account of beneficiaries

In the previous section, I suggested that McLaughlin's account of beneficiary status is grounded in the self-reproducing nature of biological systems. Given that many self-reproducing systems exist outside of the biological domain, for examples flames and hurricanes, this means that McLaughlin will be forced to ascribe these systems beneficiary status *en par* with biological systems. In order to make good on this argument, I must now present some evidence for the first half of this claim, that McLaughlin grounds beneficiary status in self-reproduction.

The primary question I am concerned with in this section is: what defines a beneficiary? Why I am the type of system that things can be good for? Why does my heart have the *function* of pumping blood, why isn't it just something that it does? Why is the circulation of oxygen *good* for me instead of something that just happens in me? The obvious answer is that I am dependent on the circulation of oxygen in order to stay alive. This move, however, raises more questions than it answers. For a start, what is it about being alive that makes it appropriate to attribute benefit and harm to living systems? And, moreover, why is it that this tree, that flower, that cat or that bird is alive – but the sun, the earth, the moon, the tides are not? What is required, if you wish to appeal to life in order to explain why some systems are appropriate subjects of benefit and harm, is a theory about what life is and how it grounds beneficiary status. To some extent, this is what McLaughlin offers with his theory of a self-reproducing system. On his account, life is a property that we attribute to self-reproducing systems – as is beneficiary status. Indeed, at some points, McLaughlin seems to imply that the property of having a good is indistinguishable from the property of being alive.[25] As such, McLaughlin could be interpreted as providing a theory of life, though I will restrict my evaluation to his theory as it applies to the property of having a good. In this section, I will examine what it is about self-

[25] McLaughlin's position is most clear on this point when he endorses the view by von Wright that: 'The question 'What kinds or species of things have a good? Is therefore broadly identical to the question 'What kinds or species of being have a life?' (von Wright quoted in McLaughlin 2001, 193, see also McLaughlin's affirmation that he endorses this view on p.195).

reproducing systems that awards them beneficiary status. McLaughlin's long and complicated argument is comprised by three key moves.

The first move McLaughlin makes is to ascribe organisms a characteristic activity, based on the Aristotelian notion of an *ergon*. Broadly speaking, a characteristic activity is 'what a system does that makes it what it is'. Typically, the characteristic activity simply *explains* why we call an object by a particular name: the reason why I call the sharp bladed objects in my top-drawer knives is because their characteristic activity is cutting. This is what a knife does that makes it what it is. By contrast, McLaughlin argues, the characteristic activity of an organism is self-reproduction. And this makes an organism what it is in a very differently way from the way cutting makes a knife what it is. To put it simply through self-reproduction, the organism *literally* makes itself what it is. It does not simply *explain* why we call something a self-reproducing system, it *justifies* why we call it a self-reproducing system.

Having thus ascribed organisms a characteristic activity, McLaughlin argues that we can identify features and properties which are *good for* the system's ability to perform its characteristic activity. In one very undemanding sense of *good for*, this seems to be true of all systems that have a characteristic activity. To distinguish it clearly, let's call this type of good a performance good. Things which aid the system in performing this function can be said to be *good for* the system's performance and things which hinder the system from performing its function can be said to be bad for the system's performance. There is, however, a more demanding sense of *good for*, which McLaughlin seems to have in mind. In this interpretation, when something is good for a system, it is taken to be good for the system's welfare. McLaughlin then argues that things can only be good for a system in this more demanding way if it is a self-reproducing system. McLaughlin's exact argument is as follows:

> 'If the characteristic activity of an organism is its self-reproduction, then 'good for the characteristic activity of X' and 'good for X' are the same. This is what makes organisms, as opposed to inanimate objects, the appropriate subjects of benefit and harm.' (2001: 203)

This brings us to the third and final move that McLaughlin makes in his account of beneficiaries. In the first step of the argument McLaughlin assigned organisms a characteristic status: self-reproduction. In the second step, McLaughlin argued that things can be good or bad for helping a system to perform its characteristic activity. From this, the third step McLaughlin makes it to argue that if some feature is good for the self-reproduction of X, it is also 'good for X'. That is, McLaughlin goes from talking about an instrumental good to an intrinsic good. The important question we must turn to here is why: why are things which are good for the self-reproduction of X good for X? The answer that McLaughlin gives is that self-reproduction is necessary for X's preservation; it is only when the preservation of a system is dependent on the performance

of its characteristic activity that things which are good for the characteristic activity of X are good for X.[26] The hidden assumption seems to be that preservation itself is good for X; if preservation is good for X and self-reproduction leads to preservation then, *ipso facto*, self-reproduction is good for X. What McLaughlin does, then, is not only to go from talking about an instrumental good to an intrinsic good, but also from talking about a performance good to a welfare good.

The account provided does justice to common sense. There is, however, a rather glaring hole in it, namely, what justifies the claim that preservation is good for X? While it certainly seems justifiable to say that a steady heartbeat is good for self-reproduction, which in turn is good for preservation, McLaughlin goes further than this, his account is committed to the claim that preservation is *good for X*. The alternative move would be to say that self-reproduction is *good for X's preservation*, without casting any judgement as to whether preservation itself is good for X. What this would do is shift the emphasis from being a good end (i.e. preservation is good for X) to being a good means to an end (i.e. self-reproduction is good for X's preservation). This would have a distinct advantage, it does not require the justification of the further claim that preservation is good for X. But it would also have a distinct disadvantage, it would undermine the ambition to ascribe beneficiary status to X. If it is not the case that self-reproduction or preservation is *good for X*, then we cannot justify the claim that the circulation of blood is *good for me*, it is simply good for preservation.

What McLaughlin's account seems to amount to is the idea that whenever we say 'the circulation of blood is good for me', this is really an elliptical way of saying that 'the circulation of blood is good for self-reproduction which is good for preservation which is good for me'. Without this final premise, none of the preceding premises ladder up to anything which is good for X. And what this premise seems to rely on is the intuitive idea that preservation is intrinsically good for X.

We have now traced McLaughlin's account of biological beneficiaries to its very foundations. In the first analysis, it seemed that in order to justify biological function ascriptions, we needed to account for their beneficiary status. Now we have seen that what we really need to justify is the claim that preservation is good. How does McLaughlin attempt to justify this claim? As far as I can make sense of it, McLaughlin

[26] In fact, this argument is formulated as a negative argument against the common tendency to ascribe beneficiary status to artifact systems. To this end, McLaughlin says: 'In artefacts, as we have seen, preservation and performance are only contingently connected… We can only really be misled into literally attributing a good to artefacts when the performance of their functions is also good for, or at least compatible with, their preservation as function bearers' (2001: 203). The implication of this seems to bet that, in organic systems, the preservation of a function bearer with the performance of its function is not contingently connected, and this demarcates biological functions from other types of causal systems.

seems to imply that existence is good for self-reproducing systems because of their self-reproducing organisation. As McLaughlin puts it,

> 'The real metaphysical cost of functional explanations lies in a commitment to the existence of entities that can stop a functional regress. In order to do this, the entities, in a sense, must have an instrumental relation to themselves… The regress of functions is stopped by turning in on itself: To say that something is good for the reproduction of a system of this kind is to say that it is good for the system' (2001: 211).

The core of the idea seems to be captured in the claim that self-reproducing systems have 'an instrumental relation to themselves'. Unlike other types of systems, self-reproducing systems are organised to maintain their own existence. But if this is the response given, this would imply that any organisationally-directed system is functional, including the solar system and the stick in the vortex, not to mention other classic counter-examples like the marble which, thrown into a bowl, pursues the goal of getting to the bottom of the bowl (Bedau 1992, 36). So it cannot be a simple matter of directed-organisation, McLaughlin must appeal to some additional criteria. And it is here that the account implicitly appeals to the goodness of preservation. It is because self-reproducing systems are organised towards the specific end of maintaining their own existence that they are beneficiaries of their activity while the marble and bowl or solar system is not. And what is it about this end that justifies singling it out? It seems like the only sensible response to give at this point is the fact that existence is good.

It doesn't seem to matter which way we approach the issue; McLaughlin's account turns on the assumption that preservation is intrinsically good. Or, put negatively, if preservation is not good for an organism, what is? As McLaughlin himself puts it: 'it is hard to see how something could be good for the repair or "re-production" of an organism in this technical sense without being good for the organism qua organism' (2001: 183).

Before moving on, it is worth returning to the issue of how McLaughlin's account treats dissipative systems such as flames and hurricanes. Recall that McLaughlin appeals to the beneficiary nature of organisms to distinguish them from simple self-reproducing systems like flames and hurricanes. This faced us with the question, what makes it the case that the organism is a beneficiary and the flame is not? We now have an answer to that question: the organism is a beneficiary because it is organised so as to maintain its own existence and existence is good. But if this is true then so too is it true that hurricanes, flames and other dissipative systems are beneficiaries. They are also organised to maintain their own existence and I can see no justification why existence should be considered good for organisms but not for them. This is why, in the previous section, I suggest McLaughlin's account fails to properly delimit the boundary between functional

and non-functional systems. This suggests that according to McLaughlin's account, flames and hurricanes are functional systems.

Internal inconsistencies in McLaughlin's account

In this paper, I have examined McLaughlin's claim that biological systems are beneficiaries. First, I tried to clarify McLaughlin's arguments as to why biological functions must be beneficial and biological systems beneficiaries. To this end, I argued that McLaughlin appeals to beneficiary status in order to evade a number of pesky counter-examples which would otherwise plague his account. Ultimately, I have argued that McLaughlin's account struggles to restrict beneficiary status to the biological domain. Because he grounds beneficiary status in self-reproduction, McLaughlin seems forced to grant beneficiary status to physical self-maintaining systems such as hurricanes and flames.

In addition, I have argued McLaughlin's approach is fundamentally problematic. It suggests that whenever biologists describe a functional trait as being good for the system they implicitly appeal to their own beliefs about what is good for a system. The intuitive and appealing stance that McLaughlin adopts is the assumption that preservation is good for living systems. This assumption is indeed so intuitive and appealing that one scarcely notices that it is just that, an assumption. It is not grounded in objective facts about biological systems. And yet it lingers underneath the surface of McLaughlin's account, acting as a primary foundation on which the rest of the account relies. This, in turn, raises a number of further consequences for McLaughlin's account.

From final causes to formal causes and back again

McLaughlin claims that he has provided an account of functions which appeals to the Aristotelian notion of formal causes, not final causes. This was important as much of the controversy connected to functions stems from their apparent connection to teleology and, in turn, final causes; final causes having been banished from science by Descartes who recognized in them an illicit appeal to psychological intentions (at that time, God's intentions). My analysis challenges McLaughlin's claim that his account rests on an appeal to formal causes not final causes.

To take the classic example, the formal cause of a house is its blueprint and its final cause is shelter. To explain the structure of the house or house-building process, we can appeal to either the final or formal cause but we do so in response to two different questions. We reference the formal cause, the blueprint, if we are asked 'What is it?' and we reference the final cause, shelter, if we are asked 'What is it for?' Following Aristotle, McLaughlin argues that, in the case of organisms, the final and formal causes can be viewed as the same thing:

‘In embryological development, for instance, the form of the mature organism is both the goal of the process of development and that for the sake of which development occurs. The questions, “What?” and “For the sake of what?” have the same answer: the organism’ (2001: 18).

In the rest of *What Functions Explains,* McLaughlin fleshes out this idea by appealing to the organism’s form: specifically, its self-reproducing structure. This provides a more detailed and compelling response to the questions ‘What?’ and ‘What for?’ The problem is that by tying his account of functions to an evaluative claim about preservation, McLaughlin sneaks final causes back into the picture. The form of the organism (self-reproduction) is not good in and of itself. To ground the goodness of the form of the organism, McLaughlin must ultimately appeal to the fact that the form is necessary for the organism’s preservation. The form is only good, is only a final end, insofar as it leads to the system’s preservation. But preservation is a final cause, not a formal cause. It seems in order to answer the question ‘What?’, McLaughlin must ultimately appeal to the question ‘What for?’

The organism does have a function

McLaughlin frequently claims that whole organisms do not have functions, only the parts of organisms have functions. His motivation for doing so is to block any claim that organisms serve an external, instrumental function. For example, to block claims such as ‘it is the function of grass to be eaten by sheep’. The impetus for this is not only to distinguish a biological account of functions, demarcating them from artifact functions, but also to block the idea that organisms are some kind of divine artifact (2001: 141-145). One consequence of my analysis is that it suggests that the organism does have a function and that function is self-reproduction.

The first problem for McLaughlin arises when he appeals to the notion of preservation. In doing so, as I argued above, he re-introduces final causes into the picture. Put simply, when McLaughlin appeals to preservation as the final end which self-reproducing systems are *for*. In doing so, he positions self-reproduction as the *means* to preservation. In turn, this implies that organisms do have a function and that function is self-reproduction.

This becomes particularly obvious when you consider the filled in elliptical statement I proposed to make sense of McLaughlin’s account. As I put it, anytime McLaughlin claims the function of the heart is to circulate blood, he is committed to something like ‘the circulation of blood is good for self-reproduction which is good for preservation which is good for me’. If this is correct, then it seems plain that McLaughlin does treat self-reproduction as a type of function. The very fact that self-reproduction is *good for* preservation implies that the relation between self-reproduction and preservation is not merely constitutional, nor is it merely cause-effect; it is a means-end relation. According to McLaughlin’s own analysis, as soon as

you start talking about one activity as being *good for* another activity, you are talking about that activity as a means to an end (2002: 133).

The second problem arises when McLaughlin describes self-reproduction as the characteristic activity of biological systems, basing this on the Aristotelian concept of an *ergon*. The difficulty is that *ergon* and function are often taken to mean the same thing. If McLaughlin wishes to claim simultaneously that organisms don't have a function but do have an *ergon*, he also needs to provide an account of where the difference between an *ergon* and a function is supposed to lie. McLaughlin fails to provide such an account. An *ergon*, McLaughlin claims, is what a thing does which makes it what it is. But so too is a function, especially when it is constituted by the notion of a feedback mechanism. In addition, the examples which McLaughlin provides of an *ergon* are all classic examples of function: the *ergon* of a sheepdog is to herd sheep, so too is its function; the *ergon* of a pruning knife is to prune, so too is its function; the *ergon* of a car is to transport, so too is its function (2001: 200-204). Without clearly spelling out what the difference between *ergon* and function is, it is hard not to interpret these as examples of functions.

What all this seems to suggest is that, despite McLaughlin's protestations to the contrary, organisms do have a function and that function is self-reproduction. Is this a problem? Does this mean that organisms might be some kind of divine artifact? The short answer is no. Just because whole organisms have functions, does not mean that they have external, instrumental functions. Nor does it justify an analogy between organisms and artifacts which might lead one to envisage organisms as some kind of divine artifact. If self-reproduction is a function, it is a very different type of function to the type of function which an artifact has. External functions benefit an external system and this external benefit explains the function bearer's existence. By contrast, the self-reproducing function of organisms does not benefit some external system, it benefits the organism itself. When I self-reproduce, I maintain my identity, which is good for me. Given that 'I', 'my identity' and 'me' all refer to the same entity, the value is not instrumental. Often it is described as being intrinsic or internal. In this way, it differs significantly from external functions. But saying that organisms have an internal function and not an external function is very different from saying that they have no function at all.

Is preservation the only good and the only goal of biological systems?

By far the most interesting question raised in this paper is the question of whether preservation is actually good. Whether we can justify this claim in a biological context, independent of the relevance it might have in medical or ethical circumstances.

This idea seems particularly problematic when you consider the consequences it has for the way we view other biological phenomena, most notably death but also (and relatedly) altruism and complex hierarchical organization. There is not room in this paper to canvas the issues in detail but broadly, the problem is that if

preservation is good, this might be taken to imply that death is bad. Indeed, at various times, McLaughlin's arguments seem to appeal as much to the intuition that ceasing to exist is bad as the idea that preservation is good.[27] The basic problem is that every biological system does die. This is a fact. Indeed, many biological systems seem *organized* to die. Going even further, in some circumstances, certain types of death are even described as functional. And it seems inarguable that death is necessary for life to exist in hierarchically nested complex systems. These are facts about biological systems which seem just as plain as the fact that they are self-reproducing. Again, we are confronted with the classic question: how do you get from a claim about facts to a claim about norms? If biological systems are organised to both live and die, in order to judge one state as good and the other as bad, we must introduce additional assumptions. Which is exactly what McLaughlin has done. But does introducing one assumption, the idea that preservation is good, also commit us to the idea that death is bad? If so, the chasm which McLaughlin is leaping over when he makes the jump from facts to norms might be slightly harder to clear than it first appeared.

Not only is preservation assumed to be the only thing which is good for biological systems, it is also their only goal. This requires that all functional biological traits must contribute to their own preservation. This, in turn, raises questions about how this approach to functions could ascribe functional status to a variety of altruistic traits.

Conclusion

In this paper, I have drawn attention to the role which the goodness of preservation plays in McLaughlin's account of biological beneficiaries. In doing so, I have argued that McLaughlin makes a number of implicit commitments which contradict a number of explicit claims that he makes. For a start, McLaughlin adamantly denies that whole organisms have biological functions. Yet, his account of beneficiary status turns on the fact that whole organisms have the function of self-reproduction. Following from this, McLaughlin claims to have provided an account of functions which appeals to formal causes in place of final causes. And yet, in order to ground beneficiary status in the system's self-reproducing organisation, McLaughlin ultimately appeals to the connection between the system's self-reproducing organisation and its preservation. But if ever there was a likely candidate for a final cause, preservation is surely a good one.

Finally, I have argued that the only way to make sense of the link which McLaughlin draws between organisation and beneficiary status is to assume that McLaughlin is appealing to the *goodness of preservation*. This suggests that McLaughlin's entire account turns on the assumption that preservation is

[27] This is particularly obvious in McLaughlin's discussion of the torpedo, whose function is to self-destruct. E.g.: 'An efficient detonator supports the characteristic activity of a torpedo. Is it good for the torpedo?' (2001: 200).

good. That is, according to McLaughlin's account, whenever biologists ascribe biological functions, they implicitly assume that those functions contribute towards the system's preservation and that preservation is good. This raises two problems for McLaughlin. First, it seems to undermine his intention to provide a *naturalised* account of beneficiary status: the idea that preservation is good isn't justified by the account of biological organisation to which he appeals. Second, the idea that self-preservation is the only good and the only goal of biological systems seems inherently problematic, when considered with respect to altruistic biological traits.

Given all this, I would conclude by arguing that McLaughlin's organisational account is inherently problematic. Furthermore, considering the similarities between McLaughlin's account and other organisational accounts, this suggests that other organisational accounts might face similar problems. In particular, it will be interesting to see by what justification other organisational accounts are any more successful in distinguishing self-reproducing biological systems from self-reproducing physical systems like flames and hurricanes. It will also be interesting to see how organisational accounts deal with biological traits that are altruistic. By definition, such traits do not contribute to their own self-reproduction but to the maintenance or reproduction of others.

ARTICLE II: THE CHALLENGE OF APOPTOSIS FOR THE ORGANISATIONAL ACCOUNT OF FUNCTIONS

Introduction

Alternatively referred to as programmed cell death, apoptosis is the process by which both healthy and unhealthy cells actively participate in their own annihilation. From the get-go, apoptosis seems to present an obvious problem case for the organisational account proposed by Mossio et al. (2009), Saborido et al. (2011), Moreno & Mossio (2015), Mossio & Saborido (2016) and Mossio & Bich (2017).[28] As they would have it, functions are the 'specific causal effects of a part or trait, which contribute to the maintenance of the organisation and, consequently, of the part itself' (Saborido et al. 2011, 584). By contrast, apoptotic traits lead to their own annihilation. Given self-annihilation is quite plainly a different end from self-maintenance, applying the organisational account to apoptotic traits seems to result with the claim that apoptotic traits are not functional.

And yet there are many good reasons for thinking that apoptotic traits are functional. Explanations of apoptosis are both normative and teleological, the hallmark characteristics of functional explanations. And biologists explicitly talk about the functions of apoptosis. It therefore seems arguable that any complete account of functions must be able to ascribe apoptotic traits a function. If the organisational account cannot ascribe apoptotic traits a function, the account is too exclusive.

This is not the first time that the organisational account has been charged with being too exclusive. Like apoptotic traits, the organisational account also has difficulties in ascribing reproductive traits a function (Delancey 2006, Saborido et al. 2011, Moreno & Mossio 2015, Artiga and Martinez 2016, Mossio & Saborido 2016). The basic problem is that reproductive traits seem to contribute to producing and maintaining a different system from the system of which they are a part. Take the example of semen; semen is produced by a parent individual but contributes to the production of an offspring individual. As such, semen does not seem to satisfy the criteria of the organisational account. It does not maintain a system which, in turn, maintains it.

[28] Here after I will simply refer to the account developed in these five papers as 'the organisational account of functions'. When I introduce other organisational accounts, I will distinguish them by referencing their authors.

The problem with apoptotic traits is similar but different. Apoptotic traits are simultaneously part of two systems. They are part of an immediate system, the cell. And they are part of an encompassing system: cells can be a unicellular organism or they can be part of a multicellular organism; in turn, organisms are themselves part of lineages and populations. Apoptotic traits contribute to the maintenance of an encompassing system at the expense of the immediate system. For this reason, they are often conceptualised as altruistic. Their existence is particularly problematic for the organisational account because the organisational account appeals to a causal loop that exists between a trait's existence and its effects in order to ground function ascriptions. This causal loop does not exist in the case where a trait annihilates itself by performing its function.

In this paper, I explore a number of different ways that the organisational account might deal with the problem of apoptotic traits. First, I explore whether it can ground the functioning of apoptotic traits in the organisation of the cell it is part of. This approach proves a quick dead-end. I then explore whether the organisational account can ground the functioning of apoptotic traits in the contribution they make to maintaining the encompassing organization which the apoptotic cells is embedded within. Such a move would require considerable theoretical development of the organisational account. But even then, I argue, it would not overcome the basic problem. The causal loop which the organisational account appeals to simply does not exist in the case of apoptotic traits. Finally, I explore whether the organisational account might deny that apoptotic traits are functional, but provide good reasons why this would be unacceptable. Given the difficulties which the organisational accounts faces ascribing apoptotic traits a function, I propose that the organisational account adopt Delancey's (2006) proposal and consider re-framing itself as a disjunctive account.

Following this introduction, in section two I introduce the organisational account of biological systems, in which the organisational account of functions is grounded. In the section three, I introduce the organisational account of functions. In section four, I sketch the basic challenge of apoptosis for the organisational account. In section five, I explore two possible ways to save the organisational account in its current form and find neither satisfactory. In section six, I argue that the organisational account should be reframed as a disjunctive account. And in section seven, I summarise and conclude the arguments presented in the paper.

The organisational account of biological systems

The organisational account of functions is grounded in the organisational account of biological systems. As such, in order to say something about the account of functions, it will first be necessary to say something about the organisational account more generally. Out of necessity, this section pays selective attention to those areas most relevant for an evaluation of apoptosis. First, I introduce the concept of self-maintenance.

Self-maintaining systems

The organisational account characterizes biological systems by their 'circular', 'looping', or 'holistic' causal structure, formalised in the concept of closure. This closure forms between entities, defined not by their physical characteristics but by their functional role in a system. Specifically, it is claimed, these entities play a constraining role on a particular biological process. This requires the fulfillment of two conditions. First, they have a causal effect on that process (they constrain the process dynamics). Second, they are not themselves altered by that same process (at least not during the time it takes for that process to be completed. There may very well be later downstream effects of that process which in turn, alters the constraint).

To see how this is meant to work, let us consider the typical example of an enzyme. I take this example not only because it is a standard example offered in the literature (Montevil & Mossio 2015, Moreno & Mossio 2015, Bich 2018) but because the apoptotic trait I take as my case study, caspase-3, is a protease enzyme. According to the organisational account, enzymes can be characterized as constraints because (I) the chemical reaction occurs much faster under the influence of the enzyme than without and (II) the configuration of the enzyme is conserved while the reaction takes place.

Building upon this notion of constraint, the organisational account argues that biological systems exhibit a closure of constraints. In simple terms, this means that all of the constraints in a biological system are mutually dependent. Take the human vascular system as an example, focusing on the action of the left ventricle and the coronary artery. The left ventricle pumps oxygenated blood into the aorta which transports it around the body, thus ensuring the flow of oxygen from the lungs to various cells around the body. The first arteries off the aorta are the coronary arteries, which transport oxygen to the heart's own cells, including the cells in the left ventricle. Thus, cardiac cells in the left ventricle are dependent on the action of the coronary artery to receive oxygen. In this way, the coronary artery acts as a constraint, channeling oxygen around the heart. At the same time, however, the cells of the coronary artery require oxygen. Like all oxygen that is transported around the body, this oxygen is pumped out of the heart via the left ventricle. What you end up with is a relationship of mutual dependence, the coronary artery is dependent on the activity of the left ventricle and the left ventricle is dependent on the activity of the coronary artery.

Obviously this is a very simplified account, the coronary artery and the left ventricle don't function in isolation. They are part of a wider vasculature system, which in turn, is part of a wider biological system: the organism. What is important to the idea of closure is that all of the components of a biological system, such as an organism, are mutually dependent, though not necessarily directly so. At a minimum, each component of a biological system must engage in at least two direct dependency relations (Montevil & Mossio 2015). In one of those relations, another component of the biological system must be directly dependent on it. Which is to say it must play the role of a constraint. In the second of those relations, it must be directly dependent on another biological component. Which is to say, it must be constrained. In identifying all of the components

which fits these criteria, you identify the boundaries of a biological system, which (following Varela 1979) they characterize as a closed system.

It is easy to see why this system is conceptualized as self-maintaining. Because of the mutual dependency that exists between constraints in a biological system, none of these constraints could exist in isolation, they depend on each other for their existence. What's more, if one constraint fails to perform its constraining role, this will have knock on effects for other constraints which directly depend on it, which will eventually lead to propagating dysfunction throughout the system. To avoid this, the system relies on various control and regulative mechanisms through which it controls the activity of its parts. In this way, it seems sensible to say that the parts maintain the whole and the whole maintains the parts, the system self-maintains.

Levels of organization

It is a core tenant of the organisational account that closure arises at different levels of biological organisation. The most straight-forward and least controversial claim made is that a unicellular organism exhibits closure (Moreno & Mossio 2015, Chp.4). In addition, they claim that both a single cell in a multicellular organism and the multicellular organism itself exhibits closure (Moreno & Mossio 2015, Chp.6). As they put it: 'it seems reasonable to hypothesise that, in most cases, multicellular systems are self-maintaining closed organisations constituted by functionally differentiated parts (groups of cells) whose constituents (the individual cells) are themselves closed systems' (Moreno & Mossio 2015, 142-3). Finally, they have claimed that closed systems also arise across biological generations, including lineages and branches of a lineage (Saborido et al. 2011, Saborido & Mossio 2016; for criticisms of this approach see Artiga & Martinez 2015).

The core insight driving these claims is the fact that biological systems exist in nested hierarchies. A cell can be simultaneously perceived as an individual system and as a part of an encompassing system. The novel claim is that both these systems can be characterized as exhibiting closure.
But this raises an immediate problem, which I will call the problem of blurry boundaries. If a cell exists inside of a multicellular system, surely it and all of its internal functioning must contribute to the maintenance of the encompassing system. If it did not, it would not (by definition) be part of the encompassing closed system. But if this is true, then there is no need to conceptualize the cell itself as a closed system. Surely all of its internal activity can be explained by reference to the maintenance of the organisation of the encompassing closed system. This seems to suggest that a multicellular organism is a single complex system, as opposed to a complex system which is itself comprised out of smaller complex systems.

Arguing against this, Montevil & Mossio (2015) have claimed that individual cells, let alone the internal parts of a cell, do not exert the necessary causal influence at the level of the organism in order to be

characterized as a constraint. At the level of the organism, constraints are constituted by populations of cells formed together into tissues and organs. Only on mass, can cells exert the causal influence necessary to be characterized as a constraint. As such, there is no mutual dependence between an individual cell (and its parts) and the multicellular organism. Montevil & Mossio claim that this allows 'their respective closures to be separated, even though they realise a nested hierarchy (the closure of the cells is nested within the closure of the encompassing system)' (2015, 190).

The organisational account of functions

The general thrust of the organisational account is to interpret functions as 'specific causal effects of a part or trait, which contribute to the maintenance of the organisation and, consequently, of the part itself' (Saborido et al. 2011, 584). Which is to say, the organisational account ascribes functional status to those entities which fulfil the role of a constraint in a closed organisation. Let's take a close look at three formal conditions which the organisational account identifies. According to the organisational account:

> 'A trait T has a function if, and only if, it exerts a constraint subject to closure in an organisation O of a given system. This definition implies the fulfilment of three different conditions:
>
> C1. T exerts a constraint that contributes to the maintenance of the organisation O;
> C2. T is maintained under some constraints of O;
> C3. O realises closure' (Moreno & Mossio 2015, 73).

Four things are emphasized in this account which are important to elaborate.

Mutual dependency between activity and existence

One of the key consequences of closure is that all the constituents of a closed system are dependent on each other for their own existence. Moreover, the existence of each constituent is dependent on the effects of its own activity. As Moreno & Mossio put it: 'the heart is there because it pumps blood (otherwise the organism and, thus the heart, would disappear), and pumping blood is a consequence of the heart's being there' (2015, 73). A trait T must continue to perform its constraining role, otherwise those traits which depend on it would fail, causing a cascade of failures throughout the system and its eventual collapse. This would include the destruction of the trait T. In this way, the trait T must maintain its own activity in order to maintain its own existence. This interdependence between activity and existence sits at the heart of the organisational account of functions. It is this which (it is claimed) to naturalise both the normativity and teleology inherent in functional explanations.

Starting with the teleological dimension, the organisational account appeals to a causal loop that arises between a trait's existence and its effects in order to naturalise the teleological dimension of function ascriptions. Because a trait's existence (and therefore its ability to perform its constraining role) depends on it performing that constraining role, we can legitimately explain the trait's activity by appealing to its effects. If the trait stopped performing that constraining role, the entire organisation and the trait would cease to exist. As such, one of the effects of the constraining role is the persistence of the trait's own existence. Given the trait must exist in order to perform its constraining role, we can explain the activity of the trait by appealing to the effects of that activity. In short, the trait causes an effect which causes the trait to continue existing causing that same effect.

With regards the normative dimension, at its core, the organisational account ascribes a kind of value to existence itself; they claim that existence has an 'intrinsic relevance' for living systems. This intrinsic relevance, they claim, determines what norms the system should follow: 'the system must behave in a specific way, otherwise it would cease to exist' (Moreno & Mossio 2015, 71). Which is to say, the heart ought to continue circulating blood because circulating blood is necessary for the continued existence of the whole organisation and thus the heart itself. Given existence has some kind of value for the system (existence is 'intrinsically relevant'), the heart ought to do that which secures its continued existence. In short, the norms of the systems are derived from the intrinsic value ('relevance') placed on existence itself.

System-maintenance as the final end of function ascriptions

As is made plain in C1 above, the function of biological traits is to maintain the closed organisation of which they are a part (and which thus maintains them). In philosophical terms, this is the final end which grounds function ascriptions[29]. But although the organisational account only identifies the final end in C1, in the real world, the effects of a constraint form a chain of means to ends. Although C1 makes no mention of these intermediate effects, they must be referenced if C1 is to be justified in a real-world situation This can be seen quite plainly when you consider a fleshed-out example which Moreno & Mossio provide:

> 'The heart has the function of pumping blood since (C1) pumping blood contributes to the maintenance of the organism by allowing blood to circulate, which in turn enables the transport of nutrients to and waste away from cells, the stabilization of body temperature and pH, and so on' (2015, 73).

[29] For an excellent discussion of the notion of final end in the context of organisational account, see McLaughlin 2001, Chp.2

This statement identifies a number of different effects which occur when the heart pumps. First, the pumping of the heart causes the circulation of blood. This is the immediate effect of the pumping blood and is the effect which we identify as the heart's function. Second, a whole range of intermediate effects are identified, largely related to metabolism. Third, a final effect is identified, the maintenance of the organism. It is this final effect which grounds function ascriptions. According to the organisational account, this is what justifies us ascribing functional status to a biological trait; at some point down the line, its effects contribute to the maintenance of the closed organisation of which it is a part, and on which its existence depends.

Functions only arise in closed organisations

The third thing emphasized in the organisational account is that biological functions only arise in the context of closure, as expressed in C3. The motivation for this condition is to distinguish genuine functional systems from trivial cases of self-maintaining systems, which we do not commonly talk about in functional terms. In particular, dissipative systems such as flames, whirlwinds and Bernard cells would otherwise pose a threat to their account (Moreno & Mossio 2015, Mossio & Bich 2017). Like cells, dissipative systems engage in an exchange with the environment so that they self-maintain by reproducing the actual material out of which they are made. What they do not do, however, is reproduce their parts. They do not do this for the simple fact that they have no parts. They are simply a single macroscopic entity. The gist of the argument, then, is that mutual dependence only arises in the context of systems that have differentiated parts which make a differentiated contribution to maintaining the system. In cases where there are no differentiated parts, there only exists one undifferentiated entity. If there is only one undifferentiated entity there cannot be mutual dependence; mutual dependence requires the involvement of at least two entities.

Function ascriptions can only be made by abstracting functional relations from time

The final point worth noting is that although the mutual dependency relations which ground function ascriptions can only be observed in time, in order to use those relations as a suitable grounding for function ascriptions, they are abstracted from time. Which is to say, there is no mention of the time intervals at which C1 and C2 are observed. The reason for this is simple. If the chain of dependencies were expressed as they occur through time, the chain would not be circular, but merely a chain. The circular regime only exists in abstraction.[30]

[30] It is worth noting that no detailed analysis has yet been offered which explains how these abstract circular relations are supposed to naturalise teleology. To the contrary, whenever the relation between organisation and teleology is discussed, it is the actual causal relations observed in time that are referenced – not the abstract causal relations.

The challenge of apoptosis

In this section, I will ask whether the organisational account of functions can ascribe functional status to apoptotic traits, taking as my case study, caspase-3. The apoptotic process is typically divided into three stages: the triggering of the apoptotic process which can happen following both external and internal signals, the execution of the apoptotic process during which the cell is effectively dismantled, and the engulfment of apoptotic bodies by phagocytosis. Caspases play a key role during the second stage of apoptosis, they actively and effectively cause the destruction of the cell from within. For this reason, they are one of the most interesting traits involved in apoptosis for they actively participate in their own destruction. And as one author descriptively put it, they do so 'in a manner reminiscent of a well planned and executed military operation' (Thornberry & Lazebnik 1998).

Caspases are typically distinguished into two groups: initiator and executioner/effector caspases (e.g. Thornberry & Lazebnik 1998, Reed 2000, Lockshin & Zakeri 2004, Elmore 2007). Initiator caspases self-activate in response to either an intrinsic apoptotic signal, such as the leakage of cytochrome-c from the mitochondria, or extrinsic signals such as the engagement of death receptors on the cell surface. When activated, these initiator caspases cleave and activate further caspases, causing a caspase-cascade, which eventually activates the executioner caspases. It is the executioner caspases which then effectively dismantle the cell. Of all the executioner caspases, caspase-3 is usually considered to be the most important. In non-apoptotic cells, CAD is coupled to its inhibitor ICAD, which prevents its release. Caspase-3 cleaves ICAD which in turn releases CAD into the cell which then degrades chromosomal DNA and causes chromatin condensation, the shrinkage which is the tell-tale morphological sign of apoptosis (Liu et al. 1997, Enari et al. 1998, Elmore 2007). In addition, caspase-3 induces the disintegration of the cell into apoptotic bodies (Sakahira et al. 1998).

Biologists typically appeal to two different downstream effects of ICAD-cleavage in order to explain the activity of caspase-3. First, explanations appeal to the immediate effect of this activity, the activation of CAD. For example, 'Caspase-3 specifically activates the endonuclease CAD… In apoptotic cells, activated caspase-3 cleaves ICAD to release CAD' (Elmore 2007, 8). Let us call this the immediate effect explanation. CAD release is the immediate effect of ICAD cleavage and so explanations of ICAD cleavage which reference CAD release are immediate effect explanations. Second, explanations appeal to the role of CAD activation in apoptotic cell death. For example, 'One role of caspases is to inactivate proteins that protect living cells from apoptosis. A clear example is the cleavage of ICAD, an inhibitor of the nuclease responsible for DNA fragmentation, CAD. (Thornberry & Lazebnik 1998, 1313). In this explanation, CAD release goes from being an end in itself to being a means to the downstream end of apoptosis. I will call this the apoptotic explanation.

Finally, biologists have identified a variety of different roles which apoptosis serves in the maintenance of the encompassing system it is part of. These can be grouped into four different types.

Developmental functions

During the embryogenesis of an organism, apoptosis has been shown to play two vital roles. First, it shapes the very form of the organism, culling away surplus cells such as the inter-digital finger space in mammals (Mori et al. 1995). Second, apoptosis generates high specificity. This is particularly important during the development of the nervous and immune systems. Both these systems arise through an overproduction of cells, followed by the removal of cells which fail to establish functional specialisations in the form of either synaptic connection or antigen specificities (Nijhawan et al. 2000, Opferman & Korsmeyer 2003).

Damage-control functions

One of the key functions of apoptosis is the elimination of cells which have sustained DNA damage, are virally infected or have suffered exposure to noxious agents (Greenhalgh 1998, Evan & Littlewood 1998). That is to say, unlike the cell death which occurs during development, this type of cell death involves the elimination of unhealthy cells which pose a risk to the organism. One of the most striking aspects of this type of apoptosis is that it can be triggered by either an internal or external pathway (Reed 2000). From outside, following recognition of infection by surrounding cells, a death signal is delivered to death receptors on the surface of the (soon to be) apoptotic cell via cell to cell contact. Alternatively, the apoptosis inducing death signal can come from within the cell itself. A variety of factors, including DNA-damage, growth factor withdrawal and calcium overload, can activate an internal pathway leading to apoptosis (Newmeyer & Ferguson-Miller 2003).

Homeostatic functions

When Kerr, Wylie & Curie (1972) first identified apoptosis as a separate biological phenomenon to necrosis, they simultaneously posited that this type of cell death, which can be observed quite plainly during embryogenesis and metamorphosis, might also contribute to homeostasis in a mature organism. It is now estimated that roughly 10 billion cells die by apoptosis every day in an adult human just to maintain a balance with cell proliferation (Renehan et al. 2001).

Fitness functions

The discovery of apoptosis in single-cell organisms was a challenge to researchers. While the function of apoptosis in multicellular organisms seems quite clear, the apoptotic cell is sacrificed for the good of the organism, it was not at all clear what the function of apoptosis was in their single-cell counterparts. As Evan & Littlewood succinctly put it, ‘For a unicellular organism, repair of damaged DNA is the only sensible way to ensure survival’ (1998, 1317). It is now accepted that apoptosis in single-cell organisms is a form of altruism, it reduces the number of offspring an organism produces but increases the number of offspring genetically-related organisms produce. Take, for example, apoptosis in unicellular chlorophytes (Durand et

al. 2011, Durand et al. 2014). In these organisms it has been shown that apoptosis can increase the fitness of surrounding cells so long as those cells are of the same species. If an apoptotic cell is surrounding by chlorophytes of a different strain, the apoptotic cell will reduce the fitness of those surrounding cells. As such, it has been hypothesized that apoptosis evolved through some kind of kin-selection mechanism and serves the function of increasing reproductive fitness at the group level.

Can the organisational account ascribe apoptotic traits a function?

The question I turn to now is whether the organisational account of functions can ascribe caspase-3 a function. According to the organisational account, caspase-3 has a function if and only if it exerts a constraint subject to closure in an organisation O of a given system. Given that caspase-3 is simultaneously a part of two closed systems, this implies that caspase-3 could contribute to maintaining either of these systems. Its function could be rooted in the maintenance of the cell or the organism. By and large, the organisational account claims that the function of intra-cellular constraints is grounded in the maintenance of the cell. But obviously this will not do in the case of apoptotic traits. Caspase-3 exerts a constraint that contributes to the destruction of the cell, and destruction seems quite plainly to be a different end from maintenance. To put it bluntly, as a system, an apoptotic cell is not self-maintaining, it is self-annihilating. As such, the organisational account cannot appeal to the role which apoptotic constraints play in the maintenance of the cell.

The next obvious move is to appeal to the role which apoptotic constraints play in the maintenance of the encompassing system the cell is part of. Intuitively, this seems a far more promising avenue for any account of functions. Apoptotic traits really do have downstream effects which contribute to the maintenance of the organism or lineage they are part of. Moreover, biologists appeal to these downstream effects all the time when they give functional explanations of the apoptotic process. And yet, it is far from obvious that the organisational account can accommodate this move.

The question I want to ask is whether the organisational account can ascribe functional status to caspase-3 by appealing to the downstream developmental, damage-controlling, homeostatic and fitness increasing functions it serves. This requires the fulfilment of three conditions:

C1. Caspase-3 exerts a constraint (ICAD cleavage) that contributes to the maintenance of the organisation the multicellular organism it is part of;
C2. Caspase-3 is maintained under some constraints of the organisation of the multicellular organism;
C3. The organisation of the multicellular organism realises closure

The first problem that raises its head here concerns C1; it is simply not true that ‘C1: Caspase-3 exerts a constraint that contributes to the maintenance of the organisation the organism’. This is not true because

caspase-3 fulfils the role of constraint at the level of the cell, not the organism. At the level of the organism, as Montevil & Mossio themselves argue, populations of cells are required to fulfil the role of constraint. A single cell, let alone a single enzyme in a single cell simply does not have the causal influence to fulfil the role of constraint (2015). As such, at the level of the organism, caspase-3 is not a constraint. To evade this charge, the organisational account will need to devise a definition of caspase-3 as some kind of population level trait. It is worth noting that even if the organisational account can succeed in this regard, this seems to result with the odd consequence that single individual instances of caspase-3 do not serve a function. This seems somewhat counter-intuitive given that biologists do actually ascribe singular instances of caspase-3 a function all the time.

The second problem which raises its head is that, even if the organisational account can reinterpret caspase-3 as a population level trait, I do not see that this particularly helps with the problem at hand. Whether caspase-3 is a singular caspase or a population of them, it still doesn't seem true that caspase-3 exerts a constraint which contributes to maintaining an organisation which, in turn, maintains it. Caspase-3 exerts a constraint which contributes to maintaining an organisation by annihilating itself. This is the nature of what it does. To put it bluntly, there is no causal loop between the existence of a *token* cause and its effects. Appealing to role caspase-3 plays in maintaining the encompassing system makes sense of the fact that functional traits must contribute (not annihilate) the system they are part of, but it still cannot make sense of the requirement that, in doing so, they contribute to their own maintenance. In short, it is simply not clear how a token caspase-3 (at any level) exhibits the mutual dependency necessary for closure.

The third problem which the organisational account will face is that this move brings us face-to-face the problem of blurry boundaries. If we tie the function of apoptotic traits to the role they play in maintaining the organism, why shouldn't we tie the functioning of every intra-cellular trait to the role they play in maintaining the organism. By what logic is the function of caspase-3 grounded in the maintenance of the organism while the function of the cell wall is grounded in the maintenance of the cell. Surely, any argument that the organisational account could put forward that grounds the functioning of caspase-3 in the maintenance of the organism (as opposed to the cell) could also be applied to the cell wall.[31]

Where exactly does the problem lie? I contend that the core problem with the organisational account lies in its committed to a unified account of functions (Saborido et al. 2011). By this approach, traits of a cell function no differently, regardless of whether they contribute to the maintenance of the cell, the multicellular organism or a cross-generational system that they are part of. This model works when the self-maintenance

[31] See Artiga & Martinez 2015 for a similar argument made about the organisational account's response to reproductive trait.

of the cell contributes to the self-maintenance of the encompassing system, which in turn contributes to the self-maintenance of the cell. In every case, a trait maintains an organisation and the organisation maintains the trait and, as such, the whole organisation maintains itself through time. Apoptotic cells challenge this because they are altruistic. They annihilate themselves for the benefit of the encompassing system in which they are embedded. The base problem is that biological systems must function differently when they are part of encompassing complex systems. This is clearly demonstrated by the devastating consequences of inhibited apoptosis. A cell does not self-annihilate, instead it proliferates and a tumour forms. By applying the same account of functions to both embedded and encompassing systems, the organisational account lacks the necessary tools for dealing with these differences.

Can the organisational account deny apoptotic traits are functional?

In this paper I argue that the organisational account put forward by Mossio et al. (2009), Saborido et al. (2011), Moreno & Mossio (2015), Mossio & Saborido (2016) and Mossio & Bich (2017) lacks the resources to ascribe apoptotic traits functional status and thus requires some revision. The force of this criticism depends on the fact that apoptotic traits are proper candidates for function ascriptions in the first place. In this section, I argue that explanations of apoptosis display the three hallmark characteristics of a functional explanation; they are non-accidental, normative and teleological. Thus, any complete account of functions must be able to ascribe apoptotic traits a function.

Indeed, it was the regularity and predictability of the apoptotic process which first distinguished it from necrotic cell death (for detailed overviews, see Lockshin & Zakeri 2004 or Potten & Wilson 2004). Cell death by necrosis is the cellular equivalent of being hit by a bus. Death is induced by external toxins, infection or trauma. The cell swells, metabolic activity stops, and the cells organelles randomly begin to disintegrate. Internal cell constituents will start to leak out of the cell, causing an inflammatory immune system reaction. By contrast, when a cell dies by apoptosis, it shrinks, isolating itself from surrounding cells. Metabolic activity continues and the organelles remain healthy long into the apoptotic process. The apoptotic cell breaks apart creating regular-sized bodies called apoptotic fragments. These apoptotic fragments are then consumed and digested by specialised cells in a process called phagocytosis. During all of this, no immune system response is mounted. In short, when cells die by apoptosis, the process is predictable, organised and regular. As biologists describe it, it is ‘controlled’, ‘regulated’, ‘orchestrated’, ‘organised’, ‘coordinated’ and distinguished by a ‘characteristic pattern’. In addition, it is an active process; as the discussion of caspases show, the cell actively participates in its own death. In these ways, apoptosis differs significantly from necrosis, which by contrast can be described as passive, messy and accidental.

The second question we must turn to is whether explanations of apoptotic traits are teleological. Here, again, the answer is yes. Consider again Elmore’s statement, ‘Caspase-3 specifically activates the endonuclease CAD… In apoptotic cells, activated caspase-3 cleaves ICAD to release CAD’ (Elmore 2007, 8). Note that it

does not change the meaning of this explanation to re-phrase it as an in order to statement, the iconic form of a teleological explanation. Caspase-3 does not just happen to cleave ICAD, releasing CAD; activated caspase-3 cleaves ICAD in order to release ICAD. That is why caspase-3 is where it is, doing what it is doing.

Finally, biologists talk about apoptotic traits in normative terms all the time. Consider the role of p53, a gene which is believed to monitor for DNA damage. Signs of DNA-damage will raise levels of p53, which in turn can either lead to cell-cycle arrest, allowing the cell time to repair itself, or in more extreme cases, the expression of apoptotic proteases, which will cause cytochrome-c to leak out of the mitochondria, triggering the apoptotic pathway. For this reason, p53 has been referred to as 'the guardian of the genome' (Lane 1992), 'the guardian of replicative normalcy' (Evan & Littlewood, 1998: 1319) and the 'policeman of oncogenes' (Efeyan 2007). It is known that p53 mutations are strongly associated with tumor formations. Now consider the following statements: 'P53 mutations are associated with adverse prognosis in many sporadic cancers' (Mantovani et al 2019: 200) and 'Tumorigenesis can occur if the system goes awry. If the p53 gene is damaged, then tumor suppression is severely reduced' (Elmore, 2007: 10). All of these terms, 'normalcy', 'adverse' 'awry', 'damaged' are normative.

In sum, the explanations which biologists give of apoptotic traits position them as non-accidental, teleological and normative. Given this, it seems fair that any complete account of functions should be able to ascribe apoptotic traits functional status. An alternative possibility would be to re-frame the organisational account as providing a stipulative theoretical definition of a particular type of function, as Millikan has done elsewhere (1984). But one of the real strengths of the organisational account is supposed to be that it represents what biologists actually mean when they ascribe a trait a function. This strength would be somewhat undermined if the organisational account were to deal with the problem of apoptotic traits by simply eliminating it.

An alternative approach: Delancey's splitting account

In the last section, I argued that it is far from obvious that the organisational account can ascribe apoptotic traits a function and that this is a problem as apoptotic traits are functional. In this section, I will propose that the organisational account should follow Delancey's (2006) proposal and consider re-framing itself as a disjunctive account. This is not a suggestion that the organisational account should adopt Delancey's entire account which is markedly different from the organisational account in many ways. Rather, I argue that the most sensible way of dealing with apoptotic traits while remaining true to what biologists actually mean when they ascribe apoptotic traits a function is to re-frame the existing organisational account as a disjunctive account.

Delancey's account was originally suggested to deal with the problem of reproductive traits, which has also plagued various versions of the organisational account (e.g. Schlosser 1998, Delancey 2006, Saborido et al. 2011, Artiga & Martinez 2015, Mossio & Saborido 2016). The basic problem is that reproductive traits contribute to producing or maintaining a different system than the one they are part of. Organizational accounts stipulate that functions of traits contribute to the self-maintenance of the system they are part of and, thereby, contribute to their own self-maintenance. This creates a causal loop between a trait's existence and its effects which is claimed to justify talking about the effect of a trait in both teleological and normative terms. Such a causal loop doesn't seem to exist in the case of reproductive traits and this seems to imply that organisational account cannot ascribe reproductive traits a function.

The problem of apoptotic traits is very similar. In both cases, the causal loop which grounds the basic organisational approach to functions is broken. The source of this break, however, is different in each case. In the case of reproductive traits, the core problem is that traits contribute to producing and maintaining a different system than the system they are part of. By contrast, apoptotic traits contribute to maintaining a system they are part of but they do so by ending their existence. Thus, in both cases, the end result is that a token trait does not contribute to maintaining its own existence, but the reasons why are quite different. While reproductive traits are not part of the system they maintain, apoptotic traits are part of the system they maintain but their effects don't loop back around to maintain them.

In response to the problem of reproductive traits, Saborido et al. (2011) and Mossio & Saborido (2016) have appealed to the existence of cross-generation systems which are maintained by reproductive traits and which, in turn, maintain them. This move allows them to treat intra-generational and cross-generational traits the same. At its core, it boils down to two ontological claims. First, they argue that the traits in question are cross-generation traits. This implies that the semen of a grandfather, father and son can all be considered time-slices of the same trait, much as we would consider my heart to be the same heart regardless of whether I was ten, twenty or thirty. Second, they argue that this cross-generation trait contributes to the maintenance of a cross-generation encompassing system, an individual that extends beyond the boundaries of individual organisms. In our example, it would minimally include the grandfather, father and son. Regardless of whether this approach works in the case of reproductive traits (See Artiga & Martinez 2015 and Garson 2017 for criticisms), I do not see that this approach helps in the case of apoptotic traits. Apoptotic traits are different from reproductive traits. However you identify caspase-3, if it is identified as a token, it is simply not true that it maintains itself through time via the effects of its own activity. The reality is quite the contrary.

The alternative proposal put forward by Delancey re-frames the organisational account as a disjunctive account. This allows for the possibility that traits can be both self-sustaining or self-reproducing. It achieves this by introducing a distinction between tokens and types. Self-sustaining traits indirectly contribute to

maintaining their own existence. The story here has nothing to do with types. There is one trait, it may be the only one of its kind that has or will ever exist for all it matters, and it contributes to maintaining a closed system which, in turn, contributes to maintaining it. On the other hand, self-reproducing traits indirectly contribute to maintaining the existence of another trait that performs the same function. This requires the existence of at least two token traits of the same type. The reproduced trait might be part of the same system as the reproducing trait, or it might be part of a different system. If the reproduced trait is part of a different system than the reproducing trait, both traits must exist within systems of the same type.

From the outset, it seems quite plausible that this approach would work for apoptotic traits. For example, let us consider again the case of caspase-3. I claim that caspase-3 has the function of cleaving ICAD. According to Delancey's account (2006, 91-92), for this to be true, means that:

ICAD is a self-reproducing teleofunction of caspase-3 iff:
For a certain time period of time T, and where $t_1 < t_2 < t_3$; and where T ranges from before t_1 to after t_3,

1. There is an instance of caspase-3, $s_1(t_1)$, that is directly necessary to establish ICAD-cleavage under circumstances c_1.
2. ICAD-cleavage(t_2) is indirectly causally necessary to establish some instance of caspase-3, $s_2(t_3)$ under circumstances c_2.
3. The causal relations between caspase-3(t_1), ICAD-cleavage(t_2), and caspase-3(t_3) are part of at least one complex system; if they are part of more than one complex system, then these are of the same type of complex system.
4. It is not the case that $s_1=s_2$.

Let us pretend that the particular case of apoptosis we are dealing with eliminates a cell which otherwise would have formed a deathly tumor in the organism. It seems quite plain that in this case, caspase-3 cleaves ICAD and in doing so contributes to establishing a further token of caspase-3 within the same complex system. If caspase-3 at t_1 had not cleaved ICAD resulting in the apoptosis of the cell, a deathly tumor would have formed, killing the organism, thus preventing the formation of future token instances of caspase-3.

The organisational account has previously rejected Delancey's (2006) disjunctive account (Saborido et al. 2011). As I interpret them, there are two reasons for this. First, Saborido et al (2011) have argued that a unified account would be more 'elegant'. But be that as it may, the unified account leads directly to the problem of blurry boundaries. According to the unified account, all traits operate for the end of self-maintenance. At the same time, many traits are simultaneously part of two self-maintaining systems. As such, there seems no principled way of stipulating which self-maintaining system they are operating for. Because of this, everything risks functioning for the self-maintenance of some higher encompassing system. Because it is a disjunctive account, Delancey's account makes some headway into solving the problem of

blurry boundaries. It allows for the fact that traits function for different ends, depending on whether they function for the immediate system they are part or for some greater encompassing system it is embedded within. In this way, while it is true that the heart is simultaneously a self-sustaining and a self-reproducing trait – it contributes both to its own self-maintenance as a token and to its self-reproduction through the establishment of further tokens – these are fundamentally two different types of function so they do not risk collapsing into each other. As I argue above, this is a major problem with the unified account suggested by Mossio et al. (2009). While the unified account may be more desirable for theoretical reasons, it is not quite clear that the biological reality it represents is as elegant as it is.

The second problem with Delancey's account concerns the way it grounds the teleological dimension of function ascriptions. As Delancey has made plain (2006), the organisational account will no longer be able to appeal to the causal loop which connects a trait's existence to its effect in order to ground function ascriptions. While you might metaphorically say that a type reproduces itself, this is and can only ever be a metaphor. When you consider the actual causal relations that exist, one particular reproduces another particular of its own kind. There is no a causal loop and without a causal loop, there is no self-maintenance. Where there is no self-maintenance, the organisational account cannot appeal to self-maintenance in order to ground function ascriptions. As far as the problem of apoptosis goes, I think this result is not only acceptable but inevitable; apoptotic traits seem testament to the fact that there exists a class of functional biological traits whose functional status must be grounded in some other way. It can't be grounded in a self-maintaining causal loop because no such loop exists.

Conclusion

In this paper, I have explored whether the organisational account proposed by Mossio et al. (2009), Saborido et al. (2011), Moreno & Mossio (2015), and Mossio & Saborido (2016) can accommodate apoptotic traits. I have argued that it cannot. In particular I have identified three basic problems with their approach.

First, on their approach, individual apoptotic traits have no function. This requires some fanciful re-interpreting of the notion of a token trait in order to accommodate apoptotic traits in their account. Second, it presents us with a problem that I have called the problem of blurry boundaries. Because the trait simultaneously contributes to maintaining multiple systems of which it is a part, how do we know which system it is functioning for? Third, in the case of apoptotic traits, it simply doesn't seem true that a causal loop exists between the activity and existence of a token trait, which is necessary to ground functional status. Whether the token trait is identified as an individual caspase-3 or a population of caspase-3, the core problem remains the same.

Following from this, I have suggested that Delancey's (2006) account is far better equipped for dealing with the problem of apoptotic traits. Unlike the account proposed in Mossio et al. (2009), Saborido et al. (2011),

Moreno & Mossio (2015), and Mossio & Saborido (2016), Delancey's (2006) account introduces a distinction between tokens and types. As such, it is able to ascribe individual apoptotic traits a function, and it doesn't introduce the problem of blurry boundaries. That said, questions remain about exactly how Delancey's account grounds the teleological dimension of function ascriptions. Give there is no causal loop between different tokens, it cannot appeal to a causal loop in order to naturalise function ascriptions. I have argued that this cannot be as problematic as proponents of the organisational account make out, there really do seem to exist a whole bunch of functional traits whose existence (as a token) is not dependent on its own activity. Apoptosis is an indisputable example of such a trait.

ARTICLE III: EXTRINSIC TELEOLOGY IN SYMBIOTIC ORGANISATION

Introduction

In recent years, a number of philosophers have advocated a new approach to biological functions, which has collectively become known as the organisational approach (Schlosser 1998, Collier 2000, McLaughlin 2001, Christensen & Bickhard 2002, Delancey 2006, Edin 2008, Mossio et al. 2009, Saborido et al. 2011, Mossio & Moreno 2015, Mossio & Saborido 2016, Mossio & Bich 2017). By this approach, a trait acquires its functional status by virtue of the way it contributes to maintaining an organized, complex system on which its existence depends.

In particular, the circular causal relation between a trait's activity and existence is claimed to naturalise the teleological dimension of functional explanations[32]. Reflecting on the type of teleology which is grounded in these relations, various authors of organizational accounts have claimed that biological systems are *intrinsically* teleological (e.g. McLaughlin 2001, 149; Delancey 2006, 80; Mossio & Bich 2017, 1090), a claim which Dan Nicholson has taken up and championed in a string of recent papers (2013, 671; 2014, 355; 2018, 141). In parallel, these authors have launched a campaign against the idea that biological systems are extrinsically teleological. Typically (though Delancey is an exception), this claim is supported by emphasizing the differences between the intrinsic teleology of organisms and the extrinsic teleology of artifacts.

Often, the matter is presented as if the concepts of intrinsic and extrinsic teleology are mutually exclusive.[33] For example, Nicholson claims that 'organisms are intrinsically purposive (in the sense that their activities

[32] In addition, this circular causal regime is claimed to naturalise the normative dimension of function ascriptions (e.g. McLaughlin 2001, p.203; Christensen & Bickhard 2002, 4; Sarborido et al. 2011, 584). See Article I in this dissertation, 'Functions and the good of preservation', for a discussion of the organizational approach to biological normativity.

[33] On this point, McLaughlin seems to shift positions throughout his book-length explication of the organizational approach (2001). At times, it reads as if he is claiming that organisms are intrinsically and not extrinsically teleological; e.g. 'the important point is that the organism, if we do attribute functions to its parts or traits, presents us with a case of internal teleology, whereas human artifacts have – thus far, at least – presented us only with cases of external teleology' (149). This is further supported by the fact that he finds reproductive traits problematic as they seem to be extrinsically teleological and this would only be problematic if extrinsic teleology is itself problematic (188, 212). But at other times, he seems to acknowledge that biological systems can be both intrinsically and extrinsically teleological. This is particularly clear in his discussion of sheep

and internal operations are ultimately directed towards the maintenance of their own organization), whereas machines are extrinsically purposive (given that their workings are geared towards fulfilling the functional ends of external agents)' (2018, 141).

While I agree that there are important differences between the teleology of artifacts and organisms, I believe that emphasizing this difference has come at a cost; it has led to a wholesale rejection of extrinsic biological teleology in any form. But biological systems exhibit extrinsically teleological relations all the time, this is particularly clear in some of the relations that arise through symbiotic interactions.

In this paper, I start by offering an interpretation of the organizational approach to biological teleology. Proponents of organizational accounts have claimed that this teleology is intrinsic but what exactly does this mean? What makes it the case that a purposive relation is intrinsically teleological, rather than extrinsically teleological? To this end, I identify three conditions which can be used to identify intrinsic teleological relations. Following this, I evaluate whether or not biological teleology is ever extrinsic. After considering a variety of symbiotic interactions, I identify two different types of extrinsically teleological relations which seem to arise through these interactions. These are competitive and cooperative teleological relations. Finally, I wrap up by considering the types of extrinsic teleology that arise in symbiotic organisation and discuss their relation to other types of teleology. My analysis suggests that the ascription of extrinsic teleology to biological systems is not as problematic as proponents of the organizational approach might fear. Extrinsic biological teleology is fundamentally different from extrinsic artifactual teleology.

A short note on the scope of this analysis

In this paper, it is not my intention to critique any particular organisational account, of which there are many. Rather, my interest lies in critiquing the claim that all biological teleology is intrinsic. To do so, I adopt the perspectives of those accounts which claim that all biological teleology is intrinsic, representing what I perceive to be the basic organisational approach. This includes the accounts proposed by Collier (2000), McLaughlin (2001), Christensen & Bickhard (2002), Edin (2008), Mossio et al. (2009), Saborido et al. (2011), Mossio & Moreno (2015), Mossio & Saborido (2016), Mossio & Bich (2017); it does not include the disjunctive accounts proposed by Schlosser (1998) and Delancey (2006). By and large, my presentation of the basic organisational approach is intended to be broad enough that it represents all of these accounts. That said, it is my intention to present the strongest and most comprehensive version of the basic organisational account. Much work has been done, particularly by Saborido et al (2011) and Mossio & Saborido (2016), to

---

dogs which seem to have both an intrinsic and extrinsic telos (200). Though note, the type of extrinsic teleology he here admits is a type of artifactual, not biological, teleology.

strengthen and widen earlier versions of the organisational account, while remaining true to their general spirit. I will adopt these extensions of the organisational account in order to give the organisational approach the best possible shot of accounting for the teleological nature of symbiotic interactions.

The intrinsic teleology of biological systems

> 'an organism is intrinsically purposive in the sense that it acts on its own behalf, towards its own ends. Its telos is internal, arising from within, and it ultimately serves no purpose other than to maintain its own organization' (Nicholson 2013, 671).

In the next section, I examine the truth of the claim that *all* biological teleology is intrinsic by considering some apparent counter-examples from studies of symbiosis. To do this, we first need a firm understanding of what intrinsic teleology is. This breaks down into two sub-questions. First, what does it mean to have *intrinsic* teleology, as opposed to extrinsic teleology? Second, what does it mean to have intrinsic *teleology*, as opposed to some other intrinsic property? After answering these questions, within an organizational approach, I will offer a definition of *intrinsic teleology*.

What is meant by *intrinsic* teleology?

Philosophers have made a number of different suggestions as to what defines the nature of intrinsic properties (Marshall & Weatherson 2018). All of these suggestions presuppose that there exists some system to whom properties can be intrinsic or extrinsic to. As such, before we can turn to the question of what it means to be intrinsic to a system, we will first need to answer the question: what does it mean to be a system. To this end, the organizational approach makes two basic claims about the nature of biological systems[34]. First, they claim that biological systems are comprised of differentiated traits that are mutually dependent on each other for their existence. Second, they claim that the system (or other traits in the system) must exert some form of control over those traits.

[34] Some organizational accounts are explicitly committed to providing a detailed ontological account of biological organization which, in turn, can naturalise function ascriptions (e.g. Collier, 2000, Christensen & Bickhard 2002, Mossio et al. 2009, Moreno & Ruiz-Mirazo 2009, Ruiz-Mirazo & Moreno 2012, Moreno & Mossio 2015). Others are primarily concerned with providing an account of function ascriptions which in turns requires that their bearers possess certain characteristics (e.g. McLaughlin 2001, Edin 2008, Delancey 2008). As such, in the following analysis, I will depend heavily on those accounts whose focus is on explicating the nature of biological systems (e.g. Christensen & Bickhard 2002, Moreno & Mossio 2015) but will cross-reference other accounts wherever possible in order to highlight the general support these two criteria receive across organizational accounts.

Mutual dependence of differentiated traits

According to the organizational approach, biological systems are constituted of functional traits that are connected to each other through a unique 'circular', 'looping', 'closed' causal structure. Drilling down, this implies that all the traits which constitute the system are mutually dependent on each other for their existence (e.g. Schlosser 1998, Christensen & Bickhard 2002, Edin 2008, Mossio et al. 2009). This does not require that all the traits of a system are directly dependent on each other, but they must be indirectly dependent on each other. This means that all traits which constitute a biological system must, at a minimum, engage in at least two dependency relations (Christensen & Bickhard 2002, Moreno & Mossio 2015). First, at least one other trait in the system must be dependent on it. Second, it must be dependent on at least one other trait in the system. The resulting organization is one in which a linear chain of cause and effect folds in on itself to create a circular web in which every trait is both cause and effect of each other. The stipulation that the traits must also be differentiated is a direct consequence of the fact that they are mutually dependent. In cases where there are no differentiated parts, there only exists one undifferentiated entity. If there is only one undifferentiated entity there cannot be mutual dependence; *mutual* dependence requires the involvement of at least two differentiated traits.[35]

This framework provides firm criteria for identifying the boundaries of a biological system. All traits of the system must both depend on and be dependent on another trait in the system, for their continued existence. The focus is on a specific organization which enables each of its constitute traits to persist through time. The framework also provides firm criteria for deciding whether a trait isn't part of a system. It excludes entities on which the system may depend but which don't depend on it. For example, plants depend on the sun for photosynthesis, but the sun doesn't depend on the plants, so the sun is not part of the plant (example taken from Christensen & Bickhard 2002). This framework also excludes entities which might depend on the system but which the system doesn't depend on. For example, certain viruses depend on the system to reproduce but are not depended on by the system for anything. Both of these examples identify one-way dependency relations.

Cross-control of relations between traits

Various organizational accounts have emphasized the fact that biological systems exert some form of control over the relations between the traits of which they are constituted (e.g. Collier & Hooker 1999, McLaughlin 2001, Edin 2008, Bich et al. 2015, Moreno & Mossio 2015). Given the inter-dependent causal structure of biological systems, these control systems are necessary for maintaining the system's persistence through time. As every trait depends on another trait performing its functional role, any internal or external

[35] See Moreno & Mossio (2015, 71-73), for an extended discussion of these issues.

perturbation which effects the performance of one trait would impact not only that trait but each trait which depends on it, which would in turn affect the traits that depend on them, potentially leading to a cascade of failures throughout the system. Control mechanisms allow the system to respond to perturbations and maintain the activity of its traits within a permissible range necessary for the system's existence.
The organizational approach identifies two different ways that such control is achieved (Bich et al. 2016). In the first, closed systems can compensate for some variation in the activity of one trait which disrupts the integrity of the system by adjusting the activity of another trait elsewhere in the system through feedback loops. For example, when an increase in glucose levels causes a compensatory increase in insulin levels (example taken from Edin 2008). In these cases, the constitutive organization of traits which defines the system does not change, though the activity of those traits will. Alternatively, biological systems can switch to a different constitutive organization in order to compensate for certain perturbations. For example, when the level of glucose in an environment drops, *E. coli* expresses a cluster of previously dormant genes which enable lactose metabolism to replace glucose metabolism. This changes the very organisation of the system, switching it between various causal organisations in response to certain triggers (example taken from Moreno & Mossio 2015).

With these two conditions, the organizational account provides clear criteria for drawing the boundaries of a system: (i) the traits which constitute a system are mutually dependent on each other for their existence (ii) relations between those traits are controlled by the system. Mapping on to these two conditions of individuality, are two corresponding ideas about what it means to be intrinsic.

Grounding intrinsic properties in biological organization
The first is a fairly traditional conception of intrinsicality, attributable to Lewis (1983). It concerns the concepts of aboutness and internality: roughly put, ascriptions of intrinsic properties are *about* the *internal* components of a system. The clearest expression of this conception comes from Marshall who puts it thus: 'a state of affairs is intrinsically about a thing *x* iff *s* describes how *x* and its parts are and how they are related to each other, as opposed to how *x* and its parts are related to other things and how other things are' (2016, 240). It is easy to see how a property would be intrinsic, in this sense, within an organizational framework: that property would have to be about traits or relations between traits that are mutually regulated and dependent. If that property refers to traits which are not mutually regulated and dependent, that property would be extrinsic.

Before moving on, it is worth noting another feature of the intrinsic/extrinsic distinction which Lewis has emphasized in his account, as it bears direct relevance on the type of teleology possessed by some symbionts. Lewis provides a number of different descriptions of the nature of intrinsic and extrinsic properties. Typically, they all finish with something like the following observation: 'an ascription of extrinsic properties to something is not entirely about that thing, though it may well be about some larger whole which includes

that thing as part' (Lewis 1983, 197). This emphasizes the fact that systems often exist in nested hierarchies. Within these hierarchies, a system will have intrinsic properties, grounded in its internal constituents, and extrinsic properties, grounded in the constituents of the larger system of which it is a part. From the perspective of the larger system, these properties will be intrinsic but from the perspective of the nested system, these properties will be extrinsic.

The second relevant conception of intrinsicality comes from Ellis (2001). He claimed that intrinsic properties are properties which a thing has independently of outside forces acting on it, where external properties are properties which a thing has in virtue of those outside forces. Ellis' conceptualization is too strict to serve as a useful model for biological systems as all biological systems are, to some extent, shaped by external forces; this is evidenced in the many one-way dependency relations with which they engage. But, it does seem to capture the general spirit of the idea that biological systems can exert control over their internal parts. They do not exhibit complete control as they are not isolated, independent systems. But they do exert some control. In this claim, we find some basis for grounding claims about intrinsicality by loosening the strictness of Ellis' claims; if a property is intrinsic to a system then it must have that property in virtue of traits which the system has some control over.

Based on the way that organizational accounts individuate biological systems, I have identified two complementary ways that intrinsic properties can be assigned to those systems. The first says that intrinsic properties must be about traits which are internal to the system. In the framework of the organizational account, they must be mutually dependent and regulated. The second says that intrinsic properties must be grounded in traits which the system has some control over.

What is meant by intrinsic *teleology*?

To say that a causal relation is teleological is to say that it operates for the fulfillment of a particular end, it is means-end orientated and the end towards which it is orientated is its telos or purpose. Moreover, in a teleological explanation, the end is referenced to explain the means; i.e. the effect is referenced to explain the cause. The problems with teleology are complex and well-worn, so I will not rehash them here.[36] Suffice to say that teleological explanations have been charged with being vitalistic, invoking backwards causation, being incompatible with mechanism and positing mentalistic powers where there are none (Mayr 1992). At the same time, it is widely recognized that functional explanations in biology are teleological (Allen & Neal 2019). To ascribe a function to a trait is to identify the end for which it is a means. To take the common examples, the hand is for grasping, the teeth for chewing, the eye for seeing and the heart for circulating

[36] For a detailed overview of the problems with teleology, see McLaughlin 2001, chap.1

blood. These are their functions, the ends for which they are the means, and the purposes for which they exist. Given this, the key challenge for accounts of biological functions has always been in providing a naturalised account of the apparent teleology of function ascriptions.

The general gist of organizational accounts of functions ascribes functions to traits in virtue of the fact that those traits contribute to the persistence of a complex organization of which they are a vital part and, in doing so, contribute to their own persistence through time. In some accounts these traits are identified as structures such as the heart (McLaughlin 2001, Delancey 2007, Mossio et al. 2009), in others they are identified as processes or events such as the heart beat (Collier 2000, Christensen & Bickhard 2002, Edin 2008) and in others again they are identified as states of structures such as the heart which beats (Schlosser 1998). For simplicities sake, I will primarily talk about traits as structures but will draw attention to those instances where it makes a real difference how those traits are defined.

To see how the account works, consider the human heart. The heart beats which causes blood to circulate. In turn, blood circulation causes nutrients to be transported to and waste from the body's cells, allowing the body's cells to be maintained, repaired and regenerated. Included in these are the cells that make up the heart itself. In this way, the heart's persistence through time is dependent on its own activity. It must continue to beat, causing blood to circulate if it is to persist through time. If the heart stops beating, the whole organization would collapse, and the heart itself would cease to exist. According to organizational accounts, this process is sufficient for naturalizing biological teleology.

Specifically, organizational accounts naturalise biological teleology by appealing to the 'closed', 'circular', 'quasi-holistic', 'cohesive' and 'looping' causal structure of biological systems.[37] To return again to the example of the human heart, it is widely accepted that the function of the heart is to circulate blood. Saying this appears to be problematic because it implies that the circulation of blood causes the heart to beat which, in turn, seems to invoke backward causation, intentional and vitalistic forces. But according to organizational accounts, the circulation of blood really does cause the heart to beat. It either has done so at previous points in time (e.g. McLaughlin 2001), or will do so at future points in time (e.g. Schlosser 1998, Christensen & Bickhard 2002), or does so in the context of an atemporal causal organization that is abstracted from time (e.g. Moreno & Mossio 2015).

[37] E.g.: 'This means that there is a certain degree of *closure* in the process, in which at least some elements (or types in this case, if we assume molecules rather than a fluid) cycle in a closed way' (Collier 2000, 283); 'autonomous systems are cohesive in the sense that they interact with the environment as a causally integrated whole' (Christensen & Bickhard 2002, 17); 'The central conceptual point for the present discussion is that self-maintaining systems can be said to instantiate what we call organizational closure, i.e. a circular causal relation between some macroscopic (or higher-level) pattern or structure and the microscopic (or lower-level) dynamics and reactions' (Mossio et al. 2009, 821).

Whichever approach is adopted, the heart beats causing blood to circulate which contributes to the maintenance of the organization as a whole, including the heart itself. In this way, the heart's existence is dependent on its own activity. On this basis, when we ask the questions 'Why does the heart beat?' or 'Why does the heart exist?' we can legitimately point to the effect of blood circulation without invoking backward causation or vitalistic/intentional forces. All we need appeal to is the actual causal organization of biological systems. Either in the past, future or abstraction, blood circulation has or will or does cause the heart to beat; just as the heart's beating has or will or does cause blood to circulate.

At the center of the account, then, is a claim that all teleological traits contribute to the maintenance of a system of which they are a part, and thus contribute to their own self-maintenance. In the traditional teleological terminology, this makes self-maintenance the final end of biological teleology[38]. The heart is not teleological because it is a means to the end of circulating blood but because the circulation of blood is means to the (final) end of self-maintenance.

What is meant by *intrinsic teleology*?

We now have some notion of what is meant by teleology; teleological relations are means-end relations in which the end explains the means and teleological traits are those entities which act as means in such a relation. According to organizational accounts, all of these means-end relations must serve the final end of self-maintenance, they must maintain a system which they are part of and on which they depend for their existence. We also have some idea of what it means to be a system; systems are constituted by differentiated and mutually dependent traits over which the system possesses some form of control. Properties which are intrinsic to such a system are grounded in traits which are intrinsic to that system.

Putting these pieces together, intrinsically teleological traits serve as means to an end where:

1. both the means and end are traits which exist as part of a closed causal system *S*,
2. the relation between that means and end is in some way controlled by the system *S*, and;
3. ultimately serves to maintain *S*'s existence.

Condition (i) concerns the traits which act as means and ends in the teleological relation, it says that both those traits must be internal to a specific type of system. In the framework of organizational accounts, they must both be mutually dependent. Condition (ii) concerns the teleological relation itself, i.e. the relation between the means and the end. It says that the system must exert some control or regulation over the

[38] For a detailed discussion of the way organizational accounts identify self-maintenance as the final end of biological traits, see Article II of this dissertation, 'The challenge of apoptosis for organizational accounts'.

relation between those traits. Condition (iii) concerns the final end which all teleological relations must serve. It says that the end in (i) must ultimately contribute to the maintenance of the system's existence (which in turn maintains the trait).

It is easy to see how the human heart is intrinsically teleological in this sense: (i) both the human heart and the ends it serves, e.g. circulation of oxygen to the liver, are part of a system of differentiated and mutually dependent traits, i.e. the human body; (ii) the circulation of blood by the heart is regulated by the system, e.g. as occurs when the heart rate is adjusted to compensate for changes in blood pressure; and (iii) the circulation of blood by the heart contributes to the maintenance of the human body. For these reasons, the teleology of the human heart can be described as intrinsic. But does this mean that *all* biological teleology is intrinsic? In the next section, we will examine a number of reasons for finding this doubtful.

The extrinsic teleology of symbionts

In this section, I discuss two different ways that symbiotic interactions involve extrinsically teleological relations. The first is a type of extrinsic teleology that arises when symbionts conjoin to form an encompassing self-maintaining system: a colony, community or supraorganism. In this circumstance, individual symbionts will possess traits that serve to maintain the symbiosis itself, rather than their own self-maintenance. The second is a type of extrinsic teleology that arises when one symbiont is co-opted by another symbiont to serve its self-maintaining organization.

Extrinsic teleology in cooperative symbioses

The question of symbiosis has previously been taken up by both Nicholson (2013) and Bich (2018). In a short footnote, Nicholson notes that the structure and behavior of organisms that live in colonies might pose some problems for the claim that biological teleology is intrinsic. It seems quite intuitive that colony-members serve ends imposed on them by the wider colony and to the benefit of the wider colony (2013, footnote 4). In a similar vein, Bich (2019) considers the functional integration which occurs in mutualistic symbioses and explores whether they undermine the notion of a closed causal organisation. How can symbionts simultaneously be closed systems and open to the kind of functional integration that arises through symbiosis? Given that intrinsic teleology is grounded in the notion of closure, if these symbioses undermine the notion of closure, so too do they undermine the notion of intrinsic teleology. Like Nicholson, Bich appeals to the idea that symbionts conjoin to form an encompassing self-maintaining organization (e.g. the colony).

In many ways, it seems an intuitive and obvious response; if the teleology of a trait cannot be explained in relation to a system's own organisation, then perhaps it is serving to maintain an encompassing organization of which the system is a part? To see how this argument is meant to work, let us consider the case of biofilms, which both Nicholson and Bich briefly reference. First, I will consider reasons for believing that

biofilm-dwelling bacteria are extrinsically teleological. Then, I will consider the argument put forward by Nicholson and Bich; that it is the biofilm as a whole which is intrinsically teleological rather than the individual bacteria which are extrinsically teleological.

Biofilms are communities of bacteria that form when those bacteria are exposed to nutritional stress. Biofilms can be composed of a single bacteria species but are more likely to be composed of a multitude of different species (Elias & Banin, 2012). The bacteria which compose a biofilm start their life as free-living independent bacteria. They then attach themselves to a surface where a process of aggregation and colonization occurs. The biofilm then enters a phase of growth during which clonal production is regulated and controlled until finally, upon maturity, the bacteria disperse (O'Toole et al. 2000, Stoodley et al. 2002). The question I am interested in, for now, is whether the *individual bacteria* that constitute a biofilm are extrinsically teleological? I will argue that they are. It is important to be clear that this is not a blanket claim that *all* the traits which distinguish biofilm-dwelling bacteria from free-living bacteria are extrinsically teleological. The behavior of bacteria living within biofilms varies so widely that they make poor candidates for sweeping blanket statements. That said, there are some general characteristic of biofilms that seem to suggest the bacteria living within them are extrinsically teleological.

Differentiation and a division of labour

One of the defining features of biofilms is that they are characterized by a division of labour amongst their bacteria constituents (Ereshefsky & Pedroso 2015). It is not clear that all of this differentiation should be considered teleological. Often times, one bacteria is simply utilizing the waste products of another bacteria and that utilization in no way explains why that waste is produced in the first place (Nadell et al. 2008). But in other cases, bacteria interact in ways that seem quite-plainly to be means-end orientated. A particularly clear example of this occurs during the aggregation and colonization phases of oral biofilms, which are formed in waves (Elias & Banin 2012). The first wave of bacterial colonizers produces adhesives for sticking to a surface, a second wave might attach to the first wave and act as a bridge for a third wave of colonizers who are unable to attach to the first wave. One example of this occurs when *S. gordonii* attaches to the tooth surface, establishing the right conditions for later colonizers including *P. gingivalis*. Importantly, this is not a passive process, *S. gordonii* has been demonstrated to actively express genes which recruit *P. gingivalis* to the multi-species biofilm (Kuboniwa et al. 2006). In this case, *S. gordonii* possesses multiple traits which serve specific functions in the recruitment of *P. gingivalis*. Similar examples abound from studies of biofilms. Some bacterial species can cause other species to produce chemicals that render the whole biofilm more resistant to bacteria (Ereshefsky & Pedroso 2015), while others have traits (e.g. the secretion of specific enzymes) that function to break down the external biofilm matrix, facilitating the biofilm's eventual dispersal (Solano et al. 2014). In all these cases, the bacteria is a means to an end external to itself and this seems to contradict the first condition for intrinsic teleology, that the means and end are internal to the system.

Control and regulation within the biofilm

A second defining feature of biofilms is the use of quorum sensing to regulate the behavior of individual community members. Quorum sensing involves the secretion and detection of specific signaling molecules, called autoinducers, which build up in the extracellular environment. It is believed that biofilm-dwelling bacteria use the level of autoinducer concentrations as a proxy for determining how dense the biofilm is. This allows them to coordinate their behavior depending on the density and composition of the biofilm. Typically, biologists studying quorum sensing refer to it as a process of regulation in which the levels of autoinducers in the external environment regulate the internal functioning of biofilm-dwelling bacteria (e.g. Miller & Bassler 2001, Nadell et al. 2008, Novik & Geisinger 2008, Ng & Bassler 2009). This suggests that a whole plethora of biofilm-specific traits are regulated by forces external to the individual bacteria, which runs contra to the second condition for intrinsic teleology, which requires that traits are controlled from within.

The production of public goods

A third feature of biofilms which biologists frequently discuss is the production of public goods which are costly for the individual to produce but benefit other cells living in the biofilm (West et al. 2006). One example of a public good would be the production of the autoinducers which facilitate quorum sensing. Recently, this has been established in *P. aeruginosa*, one of the most intensely studied and therefore best understood biofilms. In this example, it was found that production of autoinducers was costly to individual bacteria as was response to those autoinducers (West et al. 2007). My interest in these public goods should be obvious: their very existence seems evidence of the fact that biofilm-dwelling bacteria possess traits which serve ends other their own self-interest. And this contradicts the third condition for intrinsic teleology. Given the above characteristics, it seems arguable that biofilm-dwelling bacteria possess multiple traits which are extrinsically teleological. That said, there are two different ways to characterize this extrinsic teleology. On the one hand, it might be the case that individual bacteria serve the interests of other individual bacteria in the biofilm. On the other, it might be that individual bacteria serve the interests of the biofilm as a whole. It is this later route which Nicholson and Bich advocate.

The success of this route lies in its ability to establish that the biofilm is an individual in the relevant sense. Given the current knowledge gaps in biofilm functionality (see Monds & O'Toole for a review), we are a long way from developing a fully fleshed out account of biofilms as causally closed self-maintaining systems. In addition, given the heterogeneity of biofilm organization, it seems quite possible that even if some biofilms do turn out to be individuals, others will not. And even in those circumstances where a biofilm can itself be considered an individual, many of the interactions between its constituent bacteria will be better explained by reference to their intrinsic telos. But caveats aside, it is easy to see why proponents of the organisational account believe them to be promising candidates for fulfilling these criteria. Indeed, it is for the exact same reasons that I have identified them as candidates for extrinsic teleology. Consider again, the division of labour that occurs within a biofilm. In the process of identifying these divisions, we identify

differentiated and functionally dependent traits, the first condition of biological individuality. In turn the existence of regulative control mechanisms such as quorum sensing satisfies the second organizational condition of biological individuality. Taken together, these characteristics suggest that biofilms are individuals within an organizational framework.

There is, however, a point to note about all of this. The fact that a biofilm is intrinsically purposeful does not negate the fact that the individual bacteria which constitute it are extrinsically purposeful. To put this in a broader perspective, consider the criteria by which Lewis draws the line between intrinsic and extrinsic properties. As Lewis makes plain, ascriptions of properties to an individual are intrinsic to that individual only if they are wholly about that individual, properties which refer to some greater system of which the individual is a part are extrinsic properties. If the bacterium is an individual, which it is, then any ends it serves which are not under its control or for its benefit are extrinsic ends to that individual. This is a point which Nicholson explicitly recognizes in his footnote. Nicholson does not seem particularly concerned with this upshot, but this seems to be grounded in the belief that this situation is a peculiar and rare circumstance and, as such, does little to undermine the general claim that biological systems are intrinsically purposeful, not extrinsically purposeful. The exception serves to establish the rule. The problem is that these circumstances are not so peculiar and no so rare.

One of the key lessons from the explosion of microbiology research over the last twenty years has been that mutualistic symbioses are the norm, not the exception. Termites, aphids, grasshoppers, wasps, cows, mice, humans; all of these are functionally integrated closed systems which are themselves composed of functionally integrated closed systems. Consider the mealybug *Planococcus* (McCutcheon & von Dohlen 2011). Inside *Planococcus,* lives a bacterial symbiont *Tremblaya*; inside *Tremblaya* their lives another bacterial symbiont *Moranella*. All three symbionts are required to produce amino acids necessary for *Planococcus*' survival. For example, synthesis of phenylalanine starts in *Tremblaya,* the product produced is then synthesized by *Moranella* before passed back to *Tremblaya* for further synthesis before finally being passed to *Planococcus* for a final stage of processing. The production of one amino acid requires the interaction of three symbionts performing mutually dependent and differentiated functions which collectively allow the whole system to maintain itself through time. The ubiquity of these types of arrangements in nature makes plain the fact that nearly all closed systems are in fact nested hierarchies of closed systems and this nesting goes the whole way down (see Gilbert 2017 and Dupre & O'Malley 2009 for reviews).

This nesting also goes the whole way up. As the organisational account has itself stressed, ecosystems are themselves prime candidates as closed systems (Moreno & Mossio 2015). If intrinsic purpose is grounded in the concept of closure then cells, organisms, symbiotic interactions and ecosystems are all candidates for being intrinsically teleological. At the same time, because they all (except maybe the ecosystem) are

simultaneously parts of encompassing closed systems, each of these intrinsically teleological systems is also extrinsically teleological, as a part in that greater whole.

The only way that the organisational account could deny that biological systems are extrinsically teleological would be to deny that those systems are proper candidates for individuals. For example, it might be tempting to claim that neither the *Planococcus*, the *Tremblaya* or the *Moranella* are extrinsically teleological because they are not proper candidates as individuals. Rather, they are parts of one symbiotic individual constituted out of the combination of the *Planococcus*, the *Tremblaya* and the *Moranella*. There are a number of problems with this approach. First, it flies in the face of much of the organisational doctrine; many proponents of the organisational account are explicitly committed to the claim that cells in a multicellular organism, unicellular organisms, multicellular organisms, symbioses and ecosystems are all types of closed systems (Christensen & Bickhard 2002, Saborido et al. 2011, Moreno & Mossio 2015). Second, as Bich (2019) himself has argued in his discussion of symbiosis, the concept of closure does not require that a system has *complete* control over its self-maintaining organization, it only requires that it has *some* control over its organisation and therefore, its existence. It seems inarguable that nested symbionts exert some control (i.e. causal influence) over their organization, this is what identifies them as symbionts in the first place. In identifying nested symbionts as closed systems with intrinsic teleology, you identify an individual which serves an extrinsic telos at a higher level.

To surmise, I have argued that symbionts who live in nested hierarchical organisations are extrinsically teleological. Further to this, the ubiquity of these symbioses seems to suggest that nearly every organism alive serves extrinsic ends, it is the rule not the exception. This extrinsic telos does not undermine an organism's intrinsic telos but depends upon it, a system can only be extrinsically teleological if it has been marked out as an intrinsically teleological system in the first place. In addition, an organism can only be extrinsically teleological in this way if it is part of an intrinsically teleological system at a higher level of organization.

Extrinsic teleology in competitive symbioses

In the previous section, I argued that most organisms are extrinsically purposeful in virtue of the way they conjoin to form encompassing self-maintaining systems. In this section, I will discuss a very different type of symbiotic association, the interaction that occurs between a parasite and its host. Parasites are symbionts which depend on their hosts, the host, however, does not depend upon the parasite – quite the contrary, hosts are harmed by the parasitic symbiosis. They are thus very different systems from mutualistic systems in which both symbionts need each other and through their cooperation conjoin to form an encompassing self-maintaining organization. In this section, I explore the type of teleology that arises through these parasitic symbioses. Ultimately, I will argue that host symbionts serve extrinsic ends. That said, the extrinsic teleology which arises through host-parasite interactions is very different from the extrinsic teleology that arises in

nested hierarchies. This suggests that there are multiple types of extrinsic teleology found in the biological world.

To make the difference between these two types of extrinsic teleology plain, consider again the two possible explanations of the teleology of bacteria living in biofilms. The first explanation suggested that a bacterium living in a biofilm functions for the benefit of other individual bacteria living in the biofilm. The second explanation claimed that the bacterium functions for the benefit of the biofilm as a whole. The first explanation focuses on the impact which a trait has on other individual bacteria whereas the second shifts the focus on to the encompassing organization of which the bacterium is a part. Because there are good reasons to believe that biofilms are biological individuals, there are good reasons to believe that this second approach is the right way to account for the teleology of at least some of the traits which emerge in biofilms. But it seems very unlikely that this approach will work when applied to non-mutualistic symbioses. In particular, it seems unlikely it will work when we consider the way parasites can transform host behavior, physiology and morphology.

This is not to claim that all the changes hosts undergo serve extrinsic ends. Like biofilms, host-parasite interactions vary widely and some host responses undoubtedly function to lessen the impact of the parasitic infection on the host (Thomas et al. 2005). For example, when male swallows are infested by mites, their song output decreases which has a negative effect on the infesting mite, as female swallows are then less likely to mate with infected males (Moller 1991).

The traits which seem extrinsically teleological are those that contribute to the maintenance of the parasite, its lineage or branches of that lineage – often at the expense of the host. Examples of such traits are plentiful. Consider what happens when a *rhizocephalan* parasite (a type of barnacle) invades various types of brachyuran crabs (Høeg 1984, Ayaki et al. 2005, Rees & Glenner 2014). In this symbiosis, the female barnacle enters the host through its gills, makes its way through the crab before its reproductive organs erupt on the underside of the crab's abdomen, creating what is called an externa. The externa is fertilized by passing male barnacles, leading to egg production. The external egg mass mimics the crab's own egg mass in both form and position and the host crab will protect and care for the eggs as if they were their own, ventilating the eggs by shaking their abdomen and grooming them with their cleaning limbs. When the eggs are ready to spawn, the crabs also assist in their dispersal. During this process, these crabs are permanently castrated, becoming non-reproducing members of their species. Other than that, they continue to go about their ordinary lives. One of the most astounding features of this process is that hosts can be both male and female. Once infected, male crabs become behaviourally and morphologically female, they grow a larger abdomen, their gonads shrink, their fighting claws stop developing, and they display all of the nurturing behavior typical of reproducing females.

Or consider the even more dramatic case of the 'zombie ant' *Camponotus leonardi* which is invaded by the fungus *Ophiocordyceps unilateralis*. After invasion by a fungus spore, the ant is manipulated into leaving the safety of its colony, finding a leaf about 25 cm off the canopy floor, on which it clamps down in a death grip which 'functions to hold the ant in place while the fungus kills it with chemicals' (Andersen et al. 2009, 430). This positions the fungus in the ideal climate to spawn and spread its spores across the rainforest floor where it can infect other ants (Roy et al. 2006). A similar process occurs when hairworms infest crickets. The larvae of hairworms live in aquatic environments where they are consumed by mosquitos or insects which are later consumed by crickets. They then burrow through to the gut of the cricket where they will develop into a mature worm. When they are ready to reproduce, the worm causes an upregulation of Wnt proteins causing the cricket to turn 'suicidal', jumping into a body of water where the worm can release its larvae, allowing the cycle to begin again (Biron et al. 2005, Biron et al. 2006, Biron & Loxdale 2013).

In all these cases, the hosts have been shown to develop traits which serve to benefit the infecting parasite and do so, at least partly, under the parasite's control. The enlarged abdomen of the crab, the ant's death grip, and the upregulation of Wnt proteins in the cricket's brain, none of these traits serve to maintain the host. Quite the contrary, in all three cases, they cause the host considerable harm.

As I see it, there are two possible ways proponents of the organizational account could interpret these traits as intrinsically teleological. First, they could claim that the host and parasite form a closed system which itself is intrinsically purposeful. That is, they could adopt the same approach here that they adopt in the case of supraorganisms. But this won't do, the functional relations between host and parasite are a one-way street. There is no mutual dependency between the crab and the barnacle, the barnacle depends on the crab but the crab does not depend on the barnacle. If it did, the symbiosis would not be a parasitic interaction, it would be mutualistic. As Bich emphasizes in his discussion of supra-biological systems, 'Functional integration requires that subsystems contribute to one another's conditions of existence by mutually controlling their functional processes in such a way as to achieve closure' (2019, 299). This *mutual* dependence and *cross*-control simply does not occur in host-parasite interactions.

The alternative approach seems much more promising: this would claim that the enlarged abdomen, the ant's behaviour and the upregulation of Wnt proteins are all simply traits of the parasite. Or, more precisely they are traits of the parasite lineage, given that all three play a role in the maintenance of the reproductive cycle of the parasite, rather than its individual lifecycle. Given lineages are claimed to be a type of cross-generational self-maintaining system (Mosiso & Saborido 2016), they are intrinsically teleological systems which could ground the teleology of function ascriptions to these seemingly problematic traits. Let's take the ant's behaviour as our example. In this case, it is quite clear that the ant's behaviour depends on the maintenance of the fungal lineage for its production and that, in turn, this behaviour contributes to maintaining the fungal lineage. For example, the death grip ensures that the ant does not fall off the leaf post-

mortem and remains in the optimal location for spore dispersal across the canopy floor. What's more, it is quite plain that the fungal lineage exerts some control over the trait; as Hughes has demonstrated 'the height, orientation and timing of the death grip can all be precisely *controlled* by the fungus' (emphasis added, 2013, 144). Given this, it seems quite plausible that the ant's behaviour is a part of the fungal lineage. There are, however, a few consequences of adopting this approach that proponents of the organizational account will need to consider.

First, this type of argument only works if one adopts an event-based ontology of functional traits. In the above, I have specifically I identified the trait with various events, e.g. the death grip. An alternative possibility is to identify the trait with the structure that carries out that process, for example, the ant's teeth which bite into the leaf, clamping it down. This is akin to the difference between assigning the function of circulating blood to the heart or the heartbeat. But whereas, in the case of the heart, it does not make a difference whether you ascribe functional status to the heart or the heart-beat - both are maintained by the effect of blood circulation – in this case, it makes a real difference. The ant's teeth do not depend on the effect of fungal spore dispersal in the same way that the heart depends on the circulation of blood. To the contrary, the ant's teeth are harmed by its actions, as is the lineage it is part of. As such, it will not work for organisational accounts which explicitly identify traits as structures (e.g. McLaughlin 2001, Delancey 2006, Mossio et al. 2009, Saborido et al. 2011, Moreno & Mossio 2015).

To see this clearly, consider the argument which Moreno & Mossio put forward to explain the teleology of the heart. According to them, the heart has the function of circulating blood because this contributes to the maintenance of an organization (the organism) 'whose overall integrity is required for the ongoing existence of the heart itself' (Moreno & Mossio 2015, 73). A similar claim is made by Delancey when he says 'the basic insight is simple but powerful. What makes a structure or event S have a teleofunction F is that it plays a role in a complex system, such as an organism, that in turn allows and supports S, including S doing F' (2006, 82). The emphasis is on the fact that the entity which exerts the constraint (the heart) is dependent on constraints exerted by the greater organization (the organism) for its existence. This is simply not the case with the ant's teeth, the crab's abdomen or the cricket's brain; all of these structures are ultimately harmed by performing their function.

A second problem with this approach is that it makes it very hard to draw the difference between the organism and its environment. For a start, the same argument which I have just applied to hosts can also be applied to animal artifacts such as bird nests, beaver dams and spider webs. For example, the beaver dam depends on and is controlled by certain dam-building traits of the beaver, in turn, the creation of dams contributes to maintaining the beaver by increasing the beaver's ability to travel via water, reducing predation (Dawkins 1982, 200). This points to the fact that there are feedback loops between an organism

and its environment which make certain features of the environment (i.e. the beaver dam) mutually dependent on certain traits of an organism (i.e. the beaver).

According to the criteria for biological individuality identified above, these artifacts seem to be part of the organism. While this might not seem too problematic, it is worth pointing out that this argument can be extended to include the entire niche of the organism[39]. For example, the construction of dams leads to the creation of ponds which in turn attracts aquatic plants that are part of the beaver's diet, if the dam is part of the beaver lineage, this would imply that so too are the pond and aquatic plants that live in it (Laland 2004, Laland & Sterelny 2006, Wells 2015). All of these participate in a feedback loop with the lineage: they are both dependent on the lineage and the lineage is dependent on them. If the ant's behavior is part of the fungus, why shouldn't the beaver's niche be part of the beaver. By what non-arbitrary line could proponents of the organizational account draw the line between the two?[40]

A third and final consequence of this approach is that it runs contrary to most explanations which biologists offer of parasite-host relations. Consider the following statement 'This study reports the impact of infection by the mermithid nematode *Gasteromermis sp*. on the secondary sexual morphology and behaviour of its adult mayfly *host Baetis bicaudatus*' (Vance 2011, 907). The implication of this explanation, along with most others, is that the host possesses certain traits which function for the parasite. This distinction between the host and parasite is crucial for allowing biologists to describe the conflict and competition which arises between host and parasite: e.g. 'the host's body becomes the site of struggle between the parasite and its chemical products on the one hand, and any host organ defenses on the other' (Wells 2015, 559). But if the organizational account argues that all these traits are simply part of the parasite, it is hard to see how such conflict can arise. The host does not possess traits which function for the parasite, only the parasite possesses traits which function for the parasite.

In order to appreciate the problem clearly, it is worth briefly considering the debate in evolutionary biology between Dawkins' theory of the extended phenotype (Dawkins 1982) and the theory of niche construction (Lewontin 1983; Odling-Smee et al. 2003, 2013; Laland 2014). Both concern feedback loops which arise when an organism both modifies and is modified by its environment, but they describe those modifications in very different terms. According to niche construction theory, such modifications alter selection pressures on an organism in its external environment. By contrast, Dawkins claimed that a genotype can find expression

[39] I am particularly indebted to Wells (2015) for his astute critique of Dawkin's (1982) theory of the extended phenotype for this argument.

[40] It is worth noting that this appears to be a wider problem with the organizational account of biological individuality and may require a more detailed explication of the boundary between an organism and its environment.

both in the organism which is the vehicle for the genes and through structures external to that organism such as the hosts of parasites, what he called the extended phenotype. As Dawkins puts it, 'if there is a parasite within the organism, say a fluke, the genes of the parasite potentially can affect the host phenotype – the extended phenotype… they normally bend the host phenotype in a direction hostile to that favoured by the host's own genes' (2012, xii)[41]. During a parasitic infection, the host's phenotype stops serving the interests of the host genotype and instead serves the interests of the parasite genotype and does so under the control of the parasite's genotype.

The crux of Dawkin's theory is that all of the host manipulations are part of the parasite's phenotype, not the host's phenotype. This would be very similar to any claim made proponents of the organizational account that host manipulations are simply part of the parasite. But there is a crucial difference; Dawkins can appeal to the divide between genotype and phenotype, between replicator and vehicle, in order to talk about the conflict that arises between host and parasite. It makes space for the possibility that one vehicle can be bent to serve the interests of two replicators; if it is serving the interests of the parasite's genotype, then it does so by harming the host[42]. But if the traits which serve the parasite's interests are simply part of the parasite, they are no longer part of the host or host lineage and so it is difficult to see how such conflict could arise.

And this brings me to my main point; host manipulations are simultaneously part of the host and parasite. They depend on the host and, in many cases, the host will continue to depend on them. For example, the crab's abdomen continues to serve its protective function for the crab as well as its reproductive function for the barnacle. It is only the most spectacular examples where a parasite will completely destroy its host. In most host-parasite symbioses, the parasite has an interest in harming their host as little as possible; often their existence will depend on it. As Wells states it, 'the challenge of the endoparasitic way of life is to breach the host's integument but at the same time to leave it in place, so that once breached, it can work in the interests of the parasite as well as the host' (2015, 559).

In this way, most host manipulations remain a part of the host's closed causal organization and hosts continue to exert some degree of control over their functioning. For this reason, most of the changes which

[41] Similar descriptions are found in most studies of host-parasite manipulations that adopt an extended phenotype framework. E.g. 'The change of host behavior is considered to be an extended phenotype of the parasite, as it can be explained as an expression of parasite genes in the host phenotype to increase parasite fitness' (Andersen & Hughes 2012, 163); 'Parasite-induced limb malformations in amphibians offer a valuable opportunity to investigate implications of the extended phenotype and broadly quantify the effects of disease on host attributes' (Goodman & Johnson 2011, 7); 'Parasites have developed a number of fascinating strategies to infect, thrive and reproduce within their hosts [1]. Among these strategies is the manipulation of host appearance and behaviour, where the altered host traits may be regarded as the parasites' extended phenotype' (Beros et al. 2015, 1)

[42] See Turner 2004 and Wells 2015 for discussions of the ambiguities and paradoxes in Dawkin's notion of an extended phenotype.

hosts undergo when they are infected by a parasite are in fact host-mediated responses (Thomas et al. 2005). For example, the parasitic trematode Trichobilharzia ocellata can suppress egg laying in its host, the snail *Lymnea stagnalis* by excreting certain products which induce the hosts immune system to release schistosomin. In healthy snails, this release would mediate a stress response but in infected snails it functions to suppress egg laying, which is necessary for the host to fulfil its role as intermediate host (de Jong-Brink et al. 2001). In these sorts of cases, parasites rely on the normal organization and functioning of the host's immune and nervous systems but re-purpose it to serve their own ends. As a part of the host's closed causal system, they are internal to the host but the purpose they serve is extrinsic; they serve to maintain the parasitic lineage, not the hosts.

To sum, parasite manipulations of host attributes seem to be extrinsically teleological and are described as such by biologists studying host-parasite interactions. It is possible to interpret such traits within an organizational framework as intrinsically teleological by accounting for them as traits of the parasite, not the host. But doing so entails a number of problematic consequences for the organizational approach. First, this move only seems to work if the functional traits are events, not entities. Second, the same move could be applied to an organism's entire niche which seems to completely dissolve the boundaries between organism and niche. Third, there is an important sense in which the traits in question are simultaneously part of the host and the parasite; as a part of the host, the purpose they serve is extrinsic.

But this type of extrinsic teleology is a very different kind to the extrinsic teleology of nested symbionts. In this case, the extrinsic teleology of the host is not grounded in the intrinsic teleology of a greater system of which that host is a part; it is grounded in the self-maintaining organization of a system of which it is not a part. On that note, let us turn to some general reflections on the two types of extrinsic teleology I have identified in symbiotic interactions, both of which are different to the type of artifactual teleology which Nicholson (2013, 2014, 2018), McLaughlin (2001), Mossio & Bich (2017) have in mind when they claim biological systems are not extrinsically purposive.

Reflections on the nature of extrinsic teleology

In the previous section, I have identified two types of symbiotic interactions which seem to imply that various biological traits are extrinsically teleological. In this section, I will briefly discuss the type of extrinsic teleology which is proposed, its relation to intrinsic biological teleology and extrinsic artifactual teleology.

> 'While the goal of artifacts does not coincide with their own existence (the goal of a knife is not to maintain itself although, of course, it has something to do with its existence since the knife is designed for a certain use), in the case of biological systems their goal and their own existence are one and the same thing: in this

> sense, the teleology is extrinsic for the case of artifacts, and intrinsic for the case of biological systems' (Mossio & Bich, 2017)

As this quote makes plain, when McLaughlin, Nicholson, and Mossio & Bich claim that biological systems are not extrinsically teleological, they clearly have in mind a Platonic model of extrinsic teleology which roots the purposiveness of an entity in the intentions of its maker. The typical example of Platonic teleology given is an artisan and his artifacts; artisans design artifacts to serve a particular purpose and in doing so they will put them together in a way that best serves that purpose. This explains why the functioning of artifacts is means-end orientated. Because the agent is an external to the artifact, this type of teleology is usually described as extrinsic. Because it provides an explanation of how the system came into being, Platonic teleology is also creationist. And because the entire explanation turns on the intentionality of the maker, the explanation is inherently mentalistic.

By contrast, the concept of intrinsic purposiveness which these authors have in mind is rooted in an Aristotelian model of teleology. As McLaughlin sums it up, 'the goals or ends involved are those of the system itself, not those of its creator; and similarly, the valuation of a goal as a "good" is made from the perspective of the entity whose good is involved, not from that of some external agent' (2001, 17). This characterization identifies two key features which distinguishes intrinsic purpose. First, the end towards which a trait is the means arises from within the system itself. It is not imposed from the outside. Second, the fulfillment of the end benefits the system. It does not benefit an external agent. In the context of the organizational account, I have interpreted 'benefit' as a contribution to one's self-maintenance and 'arises within' as regulated mutual dependence.

Quite clearly, the type of extrinsic teleology that I have been discussing in this paper does not fit with a Platonic concept of teleology. It is neither intentional, nor creationist. As such, I have no qualms with the claim that organisms lack Platonic teleology. But I don't see why this should entail the consequence that organisms completely lack extrinsic teleology. Why should Platonic teleology be considered the only game in town? Just because biological systems do not possess the *intentional* and *creationist* extrinsic teleology of artifacts does not mean that they possess no extrinsic teleology at all. To the contrary, my discussion of the teleological relations that arise between symbionts suggests that they possess at least two different types of extrinsic teleology.

The first type of teleology, which I will call cooperative teleology, does not seem too problematic for the organizational approach. It points to the fact that the organizational account currently lacks a fully fleshed out account of how traits at one level of organization affect and are affected by traits at a higher level of organization. This paper is meant to make some headway into filling that void. I have suggested that only intrinsically teleological systems can be extrinsically teleological and that this extrinsic teleology must, in

turn, be grounded in a contribution to maintaining an intrinsically teleological system of which it is a part and on which its existence depends. The circular causal relations which grounds the teleological relation remains. But because those causal relations arise between levels of nested but individuated organisations, those relations should rightly be characterized as extrinsic.

The second type of extrinsic teleology, which I will call competitive teleology, arises when a host is manipulated by a symbiotic parasite to serve its ends. Through this process, it develops traits which are extrinsically teleological. They are imposed by a parasite to serve the interests of its lineage, rather than the lineage which the host and the trait is a part of.[43] The extrinsic teleology here arises between distinct systems where the one is not a part of the other. In order to explain the extrinsic teleology of the host, we must point to the intrinsic teleology of the parasite. In some respects, this mirrors the way that Platonic accounts of teleology explain the extrinsic teleology of artifacts by pointing to the intentionality of their makers. The difference is that what we point to here is the intrinsic teleology of the parasite and its self-maintaining organisation, not its intentionality.

It is worth noting here that this type of competitive teleology seems much more problematic for proponents of the organizational account than the hierarchical teleology identified above. In this instance, the circular causal regime which grounds teleological claims is broken. Given the organizational account of functions is also grounded in this causal regime, this seems to imply that the organizational approach will have difficulty ascribing functionality to these traits. In the end, this might motivate proponents of the organizational account to argue that these traits really are simply part of the parasite lineage. But if it does so, it will have to face the three consequences outlined above. Namely, (i) this only works if traits are defined as events rather than structures, (ii) this approach would include the entire niche as part of the lineage, and (iii) would require developing some alternative account of the competitive relations between parasite and host. For these reasons, I have suggested it is more prudent to adopt an account of extrinsic teleology which ultimately grounds these traits of the hosts in the maintenance in the organisation of the parasite or parasite lineage.

In reflection it seems that both the types of extrinsic purposiveness we have identified must ultimately be grounded in an appeal to a system that is intrinsically purposeful. This means they are ultimately grounded in an appeal to a self-maintaining organization. It is the contribution to maintaining a system which is self-maintaining which makes a relation purposeful.

---

[43] This implies another way in which hierarchical teleology can arise; between individual organisms and the lineages that they form. By my assessment, all reproductive traits would be extrinsically teleological in this way.

Conclusion

In this paper, I have provided a detailed analysis of the organizational account which describes biological systems as self-maintaining systems with their own intrinsic telos. Working within the framework of the organizational account, I identified a number of conditions which serve to make a trait intrinsic to a system. First, that trait had to be mutually dependent on other traits in the system and, second, that trait had to be subject to regulation by the system. I then identified intrinsically teleological traits as those traits which contribute to maintaining a system of which they are a part. Following this, I examined a range of symbiotic interactions to discuss the type of teleological relations that arise within them. I have argued that these interactions entail two different types of extrinsically teleological relations, cooperative teleology and competitive teleology. On the one hand, cooperative extrinsic teleology does not seem particularly problematic for the organizational account, it simply points out that the intrinsic teleology of nested hierarchies entails extrinsically teleological relations between those hierarchical levels. On the other hand, competitive teleology seems to undermine the overall organizational approach to functions. There are moves the organizational account could make to re-interpret these relations as intrinsic but these moves create as many problems as they solve. As such, I have argued that these relations are better understood as extrinsically teleological relations between competitive lineages.

RESUME

Denne afhandling undersøger biologiske systemers funktionelle natur med afsæt i den organisations-baserede tilgang til funktioner. Formålet er, gennem en kritisk diskussion af forskellige organisations-baserede teorier og deres applikation på en række biologiske træk, at kaste nyt lys over biologiske systemer, deres funktionelle natur, og hvorledes de adskiller sig fra artefakter.

I introduktionen giver jeg en kort historisk gennemgang af analogien mellem organisme og maskine og argumenterer for at den drivende kraft bag denne analogi er det faktum at både organismer og artefakter er funktionelle systemer. Herefter gennemgår jeg en række forskellige tilgange til funktioner og introducerer den organisations-baserede tilgang der både anlægges og kritiseres i de artikler der udgør denne afhandling. Dernæst giver jeg en kort sammenfatning af disse artikler og drager en række yderligere konsekvenser af de argumenter de indeholder; jeg knytter forbindelsen mellem artiklerne og bestemmer forskningsresultaternes betydning for analogien mellem organisme og maskine.

I første artikel argumenterer jeg for at McLaughlins (2001) organisations-baserede teori om funktioner hviler på en værdidom der bestemmer overlevelse som det eneste gode og eneste mål for biologiske systemer, og at denne værdidom er helt central for McLaughlins tilgang til teleologi og normativitet. Dette rejser spørgsmålet hvorvidt McLaughlins tilgang med rette kan hævde at være naturalistisk. Samtidig formår den ikke at afgrænse teleologi og normativitet til det biologiske domæne.

I anden artikel undersøger jeg om den organisations-baserede teori der foreslås af Mossio et al. (2009) er i stand til at tilskrive apoptotiske træk en funktion, og argumenterer for et negativt svar. Som løsning foreslår jeg at denne organisations-baserede teori kan inkorporere den disjunktive tilgang forslået af Delancey (2006). Det åbner mulighed for at biologiske træk kan være både selvopretholdende og selvreproducerende.

I tredje artikel undersøger jeg teleologiske relationer der opstår som følge af forskellige former for biologisk interaktion og argumenterer mod den ofte fremførte påstand at al biologisk teleologi er intrinsisk. I stedet argumenterer jeg for at mange teleologiske relationer i biologiske systemer er ekstrinsiske fordi de opstår som følge af samarbejde og konkurrence. Gennem en analyse af den form for ekstrinsisk teleologi der opstår gennem symbiotiske interaktioner, argumenterer jeg for at denne er fundamentalt forskellig fra den form for ekstrinsisk teleologi artefakter besidder.

BIBLIOGRPAHY

Achinstein P (1983) The Nature of Explanation. Oxford University Press, Oxford

Allen C, Bekoff M, Lauder G (1998) “Introduction” in Allen C, Bekoff M, Lauder G (eds.) Nature’s Purposes. The MIT Press, Massachusetts, pp 1-25

Allen C, Neal J (2019) “Teleological Notions in Biology.” In Zalta ED (ed.) The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/teleology-biology/

Amundson R, Lauder G (1994) “Function without Purpose: The Uses of Causal Role Function in Evolutionary Biology.” Biology & Philosophy 9(4): 443–469. https://doi.org/10.1007/BF00850375

Andersen SB, Gerritsma S, Yusah K, Mayntz D, Hywel-Jones N, Billen J, Boomsma J, Hughes D (2009) “The Life of a Dead Ant: The Expression of an Adaptive Extended Phenotype.” The American naturalist 174: 424-33. https://doi.org.10.1086/603640

Andersen SB, Hughes D (2012) “Host specificity of parasite manipulation: Zombie ant death location in Thailand vs. Brazil.” Communicative & Integrative Biology 5(2): 163-165. https://doi.org. 10.4161/cib.18712

Artiga M, Martínez M (2015) “The organizational account of function is an etiological account of function.” Acta Biotheoretica 64:105-17. https://doi.org/10.1007/s10441-015-9256-x

Ayaki T, Kawauchino Y, Nishimura C, Ishibashi H, Arizono K (2005) “Sexual disruption in the freshwater crab (Grothelphusa dehaani).” Integrative and Comparative Biology 45: 39–42. https:doi.org.10.1093/icb/45.1.39

Ayala F (1970) “Teleological Explanations in Evolutionary Biology.” Philosophy of Science 31(1): 1-15. https://doi.org/10.1086/288276

Ayala F (1999) “Adaptation & Novelty: Teleological Explanations in Evolutionary Biology.” History and Philosophy of the Life Sciences 21(1): 3-33. https://escholarship.org/uc/item/26s4355t

Barandiaran X, Moreno A (2008) “Adaptivity: From Metabolism to Behavior.” Adaptive Behavior 16(5): 325–344. https://doi.org/10.1177/1059712308093868

Bechtel W (2007) “Biological mechanisms: organized to maintain autonomy.” In Boogerd F, Bruggeman F, Hofmeyr JH, Westerhoff HV (eds.) Systems biology, Philosophical Foundations. Elsevier, Amsterdam, pp 269–302

Bedau M (1992) “Where’s the good in teleology?” Philosophy and Phenomenological Research 52(4): 781-806. https://doi.org/10.2307/2107911

Beros S, Jongepier E, Hagemeier F, Foitzik S (2015) “The parasite's long arm: a tapeworm parasite induces behavioural changes in uninfected group members of its social host.” Proceedings of the Royal Society B: Biological Sciences 282(1819). https://doi.org/10.1098/rspb.2015.1473

Bich L, Moreno A (2016) “The role of regulation in the origin and synthetic modelling of minimal cognition.” BioSystems 148: 12-21. https://doi.org/10.2307/210791110.1016/j.biosystems.2015.08.002

Bich L, Mossio M, Ruiz-Mirazo K, Moreno A (2016) “Biological regulation: controlling the system from within." Biology and Philosophy 31: 237-265. https://doi.org/10.2307/210791110.1007/s10539-015-9497-8

Bich L (2018) “Robustness and autonomy in biological systems: how regulatory mechanisms enable functional integration, complexity and minimal cognition through the action of second-order control constraints.” In Bertolaso M, Caianiello S, Serrelli E (eds.) Biological Robustness. Springer, New York pp 123-147. https://doig.org.10.1007/978-3-030-01198-7_6

Bich L (2019) “The Problem of Functional Boundaries in Prebiotic and Inter-Biological Systems.” In Minati G, Abram M, Pessa E (eds.) Systemics of Incompleteness and Quasi-Systems. Springer, New York pp 295-302

Bickhard MH (2000) “Autonomy, function, and representation.” Communication and Cognition-Artificial Intelligence 17(3-4): 111-131.

Bigelow J, Pargetter R (1987) “Functions.” The Journal of Philosophy 84(4): 181–196. https://doi:10.2307/2027157

Biron DG, Marché L, Ponton F, Loxdale HD, Galéotti N, Renault L, Joly C, Thomas F (2005) “Behavioural manipulation in a grasshopper harbouring hairworm: a proteomics approach.” Proceedings of the Royal Society B: Biological Sciences 272: 2117–2126. https://doi.org/10.1098/rspb.2005.3213

Biron DG, Ponton F, Marché L, Galeotti N, Renault L, Demey-Thomas E, Poncet J, Brown SP, Jouin, P, Thomas F (2006) “‘Suicide’ of crickets harbouring hairworms: a proteomics investigation.” Insect Molecular Biology 15: 731–742. https://doi.org. 10.1111/j.1365-2583.2006.00671.x

Biron D, Loxdale H (2013) “Host-parasite molecular cross-talk during the manipulative process of a host by its parasite.” The Journal of experimental biology 216: 148-60. https://doi.org.10.1242/jeb.073825.

Boorse C (1976) “Wright on Functions.” The Philosophical Review 85(1): 70–86. https://doi.org/10.2307/2184255

Borisenko GG, Matsura T, Liu SX, Tyurin VA, Jianfei J, Serinkan FB, Kagan VE (2003) “Macrophage recognition of externalized phosphatidylserine and phagocytosis of apoptotic Jurkat cells--existence of a threshold.” Archives of Biochemistry and Biophysics 413(1): 41-52. https://doi.org/10.1016/S0003-9861(03)00083-3

Brito C & Marques V (2014) “The Rise and Fall of the Machine Metaphor: Organizational Similarities and Differences Between Machines and Living Beings.” The Notion of Organism, Historical and Conceptual Approaches XLIII, N.1-4: 77-111.

Christensen WD, Bickhard MH (2002) “The process dynamics of normative function.” The Monist 85: 3-28. https://doi.org/10.5840/monist20028516

Cobb M (2013) “1953: When genes became information.” Cell 153: 503-506. https://doi.org/10.1016/j.cell.2013.04.012

Collier JD, Hooker CA (1999) “Complexly Organised Dynamical Systems.” Open Systems & Information Dynamics 6: 241–302. https://doig.org.10.1023/A:1009662321079

Collier JD (2000) “Autonomy and process closure as the basis for functionality.” In: Chandler JLR, van der Vijver G (eds.) Closure: emergent organisations and their dynamics. Annals of the New York Academy of Sciences, New York, pp 280–290

Craver CF (2001) “Role Functions, Mechanisms, and Hierarchy.” Philosophy of Science, 68(1): 53–74. https://doi.org/10.1086/392866

Cummins R (1975) “Functional Analysis.” The Journal of Philosophy 72(20): 741–765. https://doi.org/10.2307/2024640

Cummins R (1983) The Nature of Psychological Explanation. MIT Press, Massachusetts.

D'Oro G, Overgaard S (2017) “Introduction.” In D'Oro G, Overgaard S (eds.) The Cambridge Companion to Philosophical Methodology. Cambridge University Press, Cambridge, pp 1-10

Darwin C (1859) On the origin of species by means of natural selection or the Preservation of favoured races in the struggle for life. John Murray, London.

Darwin F (1887) Life and Letters of Charles Darwin. John Murray, London.

Davidson D (1987) “Knowing one’s own Mind.” Proceedings and Addresses of the American Philosophical Association 60(3): 441-458. https://doi.org/10.2307/3131782

Davies, PS (2001) Norms of Nature: Naturalism and the Nature of Functions. MIT Press, Massachusetts.

Dawkins R (1976) The Selfish Gene. Oxford University Press, Oxford.

Dawkins R (1982) The Extended Phenotype. Oxford University Press, Oxford.

Dawkins R (1986) The Blind Watchmaker. Norton & Company Inc, New York.

de Jong-Brink M, Bergamin-Sassen M, Soto M, (2001) “Multiple strategies of schistosomes to meet their requirements in the intermediate snail host.” Parasitology 123: 129–141. https://doi.org/10.1017/S0031182001008149

Delancey CS (2006) “Ontology and teleofunctions: A defense and revision of the systematic account of teleological explanation.” Synthese 150: 69-98. https://doi.org/10.1007/s11229-004-6257-8

Dickinson H (1939) A short history of the steam engine. Cambridge University Press.

Dupré J & O'Malley M (2009) “Varieties of Living Things: Life At The Intersection of Lineage And Metabolism.” Philosophy & Theory in Biology 1(201306): 1-25 https://doi.org/10.3998/ptb.6959004.0001.003.

Durand PM, Rashidi A, Michod RE (2011) “How an organism dies affects the fitness of its neighbors.” The American Naturalist 177: 224–232. https://doi.org/10.1086/657686

Durand P, Choudhury R, Rashidi A, Michod R (2014) “Programmed death in a unicellular organism has species-specific fitness effects.” Biology Letters 10(2). https://doi.org/10.1098/rsbl.2013.1088

Edin B (2008) “Assigning biological functions: making sense of causal chains.” Synthese 161: 203–218. https://doi.org/10.1007/s11229-007-9160-2

Efeyan A, Serrano M (2007) “p53: Guardian of the Genome and Policeman of the Oncogenes.” Cell Cycle 6(9): 1006-1010. https://doi.org/10.4161/cc.6.9.4211

Elias S, Banin E (2012) “Multi-species biofilms: Living with friendly neighbors.” FEMS microbiology reviews 36(5): 990-1004. https://doi.org10.1111/j.1574-6976.2012.00325.x.

Ellis B (2001) Scientific Essentialism. Cambridge University Press, Cambridge

Elmore S (2007) “Apoptosis: a review of programmed cell death.” Toxicologic Pathology 35(4): 495-516. https://doi.org/10.1080/01926230701320337

Enari M, Sakahira H, Yokoyama H, Okawa K, Iwamatsu A, Nagata S (1998) “A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD.” Nature 391:43–50. https://doi.org/10.1038/34112

Ereshefsky M, Pedroso M (2015) “Rethinking evolutionary individuality.” PNAS 112(33): 10126-10132. https://doi.org/10.1073/pnas.1421377112

Evan G, Littlewood T (1998) “A matter of life and cell death.” Science 281: 1317-1322. https://doi.org/10.1126/science.281.5381.1317

Fadok VA, Voelker DR, Campbell PA, Cohen JJ, Bratton DL, Henseon PM (1992) “Exposure of phosphatidylserine on the surface of apoptotic lymphocytes triggers specific recognition and removal by macrophages.” The Journal of Immunology 148(7): 2207-2216.

French RK (1994) William Harvey’s Natural Philosophy. Cambridge University Press, Cambridge.

Garson J (2016) A Critical Overview of Biological Functions. Springer, Dordrecht.

Garson J (2017) “Against Organisational Functions.” Philosophy of Science 84(5):1093-1103. https://doi.org/10.1086/694009

Gilbert SF (2017) “Holobiont by Birth: Multilineage Individuals as the Concretion of Cooperative Processes.” In Tsing A, Swanson H, Gan E, Bubandt N (eds.) Ghosts and Monsters of the Anthropocene. University of Minnesota Press, Minnesota, pp 77-89

Godfrey-Smith P (1994) “A Modern History Theory of Functions.” Noûs, 28(3): 344–362. https://doi.org/10.2307/2216063

Goodman BA, Johnson PTJ (2011) “Disease and the Extended Phenotype: Parasites Control Host Performance and Survival through Induced Changes in Body Plan.” PLOS One 6(5): e20193. https://doi.org/10.1371/journal.pone.0020193

Greenhalgh DG (1998) “The role of apoptosis in wound healing.” The International Journal of Biochemistry and Cell Biology 30(9): 1019-1030. https://doi.org/10.1016/S1357-2725(98)00058-2

Griffiths P (1993) “Functional Analysis and Proper Functions.” The British Journal for the Philosophy of Science 44(3): 409–422. https://doi.org/10.1093/bjps/44.3.409

Glock H (2017) “Impure Conceptual Analysis.” In D'Oro G, Overgaard S (eds.) The Cambridge Companion to Philosophical Methodology. Cambridge University Press, Cambridge, pp 77-100

Hardcastle VG (2002) “On the normativity of functions” in Ariew A, Cummins C, Perlman M (eds.) Functions: New essays in the philosophy of psychology and biology. Oxford University Press, Oxford, pp 144-156

Harvey W (1616/1928) Exercitatio anatomica de motu cordis et sanguinis in animalibus with an English translation and annotations by Chauncey D. Leake. Charles C. Thomas, Maryland.

Hempel CG, Oppenheim P (1948) “Studies in the Logic of Explanation.” Philosophy of Science 15(2): 135-175. www.jstor.org/stable/185169

Hempel CG (1959/1965) “The logic of functional analysis” In Hempel CG, Aspects of scientific explanation. The Free Press, New York pp 297–330; (reprinted from Gross L (ed.) Symposium on sociological theory. Harper & Row, New York)

Holm SH (2014) “Disease, Dysfunction, and Synthetic Biology.” Journal of Medicine and Philosophy, 39(4): 329-345. https://doi.org/10.1093/jmp/jhu025

Holt L (2018) “Man As Machine.” Important Magazine, Copenhagen

Høeg J (1995) “The biology and life cycle of the Rhizocephala (Cirripedia).” Journal of the Marine Biological Association of the United Kingdom 75: 517-550. https://doi.org.10.1017/S0025315400038996

Johansen TK (2004) “Plato's Natural Philosophy: A Study of the Timaeus Critias”. Cambridge University Press, Cambridge

Jouanna J (2012) “The legacy of the Hippocratic Treatsie The Nature of Man: The Theory of the Four Humours.” In Van der Eijk P (ed.) Greek Medicine from Hippocrates to Galen: Selected Papers. Brill, Leiden, pp 335-360

Kerr JF, Wyllie AH, Currie AR (1972) “Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics.” British Journal of Cancer 26(4): 239-57. https://doi.org/10.1038/bjc.1972.33

Kornblith H (2017) “A Naturalistic Methodology.” In D'Oro G, Overgaard S (eds.) The Cambridge Companion to Philosophical Methodology. Cambridge University Press, Cambridge, pp 141-160

Kuboniwa M, Tribble G, James C, Kilic A, Tao L, Herzberg M, Shizukuishi S, Lamont R (2006) “Streptococcus gordonii utilizes several distinct gene functions to recruit Porphyromonas gingivalis into a mixed community.” Molecular microbiology 60: 121-39. https://doi.org.10.1111/j.1365-2958.2006.05099.x.

Laland K (2004) “Extending the extended phenotype.” Biology and Philosophy 19: 313–325. https://doi.org.10.1023/B:BIPH.0000036113.38737.d8

Laland K, Sterelny K (2006) “Seven reasons (not) to neglect niche construction.” Evolution 60: 1751–1762. https://doi.org.10.1111/j.0014-3820.2006.tb00520.x

Laland K (2014) “On evolutionary causes and evolutionary processes.” Behavioural Processes 117: 97-104. https://doi.org/10.1016/j.beproc.2014.05.008

Lane DP (1992) “p53, guardian of the genome.” Nature 358:15–16. https://doi.org/10.1038/358015a0

Lennox J (1993) “Darwin Was a Teleologist.” Biology & Philosophy 8(4): 409–421. https://doi.org/10.1007/BF00857687

Lennox J (2010) “The Darwin/Gray Correspondence 1857–1869: An Intelligent Discussion about Chance and Design.” Perspectives on Science 18(4): 456–479. https://doi.org10.1162/POSC_a_00018

Lewens T (2004) Organisms and artifacts: design in nature and elsewhere. MIT Press, Massachusetts.

Lewis D (1983) “Extrinsic Properties.” Philosophical Studies 44: 197–200. https://doi.org.10.1007/BF00354100

Lewontin RC (1983) “Gene, organism and environment.” In Bendall (ed.) Evolution from molecules to men. Cambridge University Press, Cambridge, pp 273-286

Liu Z, Zou H, Slaughter C, Wang X (1997) “DFF, a Heterodimeric Protein That Functions Downstream of Caspase-3 to Trigger DNA Fragmentation during Apoptosis.” Cell 89(2): 175-184. https://doi.org/10.1016/S0092-8674(00)80197-X

Lockshin RA, Zakeri Z (2004) “Introduction.” In Locksin RA, Zakeri Z (eds.) When Cells Die II. John Wiley & Sons, New Jersey, pp 3-23

Mantovani F, Coolavin L, Del Sal G (2019) “Mutant p53 as a guardian of the cancer cell.” Cell Death & Differentiation 26: 199–212. https://doi.org/10.1038/s41418-018-0246-9

Marshall D (2016) “Varieties of Intrinsicality.” Philosophy and Phenomenological Research 92: 237–263. https://doi.org/10.1111/phpr.12156

Matthen M (1997) “Teleology and the product analogy.” Australiasian Journal of Philosophy 75: 21-37. https://doi.org/10.1080/00048409712347651

Maturana H & Varela F (1973/1980) “Autopoiesis: the organization of the living.” Reidel Publishing, Dordrecht.

Mayr E (1974) “Teleological and Teleonomic, a New Analysis.” In Cohen RS Wartofsky MW (eds.) Methodological and Historical Essays in the Natural and Social Sciences. Springer, Dordrecht. https://doi.org/10.1007/978-94-010-2128-9_6

Mayr E (1992) “The idea of teleology.” Journal of the History of Ideas 53(1): 117-135. https://doi.org/10.2307/2709913

Mayr E (1993) “Proximate and Ultimate Causations.” Biology and Philosophy 8: 95-98. https://doi.org/10.1007/BF00868508

McCutcheon JP, von Dohlen CD (2011) “An interdependent metabolic patchwork in the nested symbiosis of mealybugs.” Current Biology 16: 1366-72. https://doi.org.10.1016/j.cub.2011.06.051

McLaughlin P (2001) What Functions Explain: Functional Explanation and Self- Reproducing Systems. Cambridge University Press, Cambridge

McLaughlin P (2002) “On Having a Function and Having a Good.” Analyse und Kritik 24: 130-143.

McLaughlin P (2009) “Functions and Norms.” In: Krohs U, Kroes P (eds.) Functions in Biological and Artificial Worlds. MIT Press, Cambridge, pp 93-102

McReynolds P (1980) “The Clock Metaphor in the History of Psychology.” In Nickles T (ed.) Scientific Discoveries: Case Studies. Springer, Dordrecht, pp 97-112

Militello G, Moreno A (2018) “Structural and Organisational Conditions for Being a Machine.” Biology & Philosophy 33(5): 1-25. https://doi.org/10.1007/s10539-018-9645-z.

Miller M, Bassler B (2001) “Quorum sensing in bacteria.” Annual Review of Microbiology 55: 165 199. https://doig.org/10.1146/annurev.micro.55.1.165

Millikan R (1984) Language, Thought, and Other Biological Categories: New Foundations for Realism. MIT Press, Massachusetts

Millikan R (1989) “In Defense of Proper Functions.” Philosophy of Science 56(2): 288–302. https://doi.org/10.1086/289488

Mitchell S (1995) “Function, Fitness and Disposition.” Biology and Philosophy 10: 39-54. https://doi.org/10.1007/BF00851986

Moller AP (1991) “The preening activity of swallows, Hirundo rustica, in relation to experimentally manipulated loads of haematophagous mites.” Animal Behaviour 42(2): 251-260. https://doi.org/10.1016/S0003-3472(05)80556-1

Monds R & O'Toole G (2008) “The development model of microbial biofilms: ten years of a paradigm up for review.” Trends in microbiology 17: 73-87. https://doi.org/10.1016/j.tim.2008.11.001

Montévil M, Mossio M (2015) “Biological organisation as closure of constraints.” Journal of Theoretical Biology 372: 179–191. https://doi.org/10.1016/j.jtbi.2015.02.029

Moreno A, Mossio M (2015) Biological Autonomy: A Philosophical and Theoretical Inquiry. Springer, Dordrecht.

Moreno A, Ruiz-Mirazo K (2009) “The problem of the emergence of functional diversity in prebiotic evolution.” Biology and Philosophy 24(5): 585–605. https://doi.org/10.1007/s10539-009-9178-6

Mori C, Nakamura N, Kimura S, Ieie H, Takigawa T, Shiota K (1995) “Programmed cell death in the interdigital tissue of the fetal mouse limb is apoptosis with DNA fragmentation.” The Anatomical Record 242: 103-110. https://doi.org/10.1002/ar.1092420114

Mossio M, Saborido C, Moreno A (2009) “An Organizational Account of Biological Functions.” British Journal for The Philosophy of Science 60: 813-841. https://doi.org/10.1093/bjps/axp036.

Mossio M, Saborido C (2016) “Functions, organization, and etiology: A reply to Artiga and Martinez.” Acta Biotheoretica 64: 263-75. https://doi.org/10.1007/s10441-016-9283-2

Mossio M, Bich L (2017) “What makes biological organisation teleological?” Synthese 194(4), 1089–1114. https://doi.org/10.1007/s11229-014-0594-z

Mori C, Nakamura N, Kimura S, Irie H, Takigawa T, Shiota K (1995) “Programmed cell death in the interdigital tissue of the fetal mouse limb is apoptosis with DNA fragmentation.” The Anatomical Record 242:103-110. https://doi.org/10.1002/ar.1092420114

Nadell C, Xavier J, Foster K (2009) “The sociobiology of biofilms.” FEMS Microbiology Reviews 33: 206-24. https://doi.org/10.1111/j.1574-6976.2008.00150.x.

Nagel E (1953) “Teleological Explanation and Teleological Systems.” In Ratner S (ed.) Vision and Action: Essays in Honour of Horace M. Kallen on His 70th Birthday. Rutgers University Press, New Brunswick, pp 537–558

Nagel E (1961) The Structure of Science. Harcourt Brace, New York

Neander K (1991) “Functions as Selected Effects: The Conceptual Analyst’s Defense.” Philosophy of Science 58(2): 168–184. https://doi.org/10.1086/289610

Newmeyer DD, Ferguson-Miller S (2003) “Mitochondria: releasing power for life and unleashing the machineries of death.” Cell 112(4): 481-90. https://doi.org/10.1016/S0092-8674(03)00116-8

Ng WL, Bassler B (2009) “Bacterial Quorum-Sensing Network Architectures.” Annual review of genetics 43: 197-222. https://doi.org/10.1146/annurev-genet-102108-134304

Nicholson D (2012) “The Concept of Mechanism in Biology.” Studies in History and Philosophy of Science Part C 43(1): 152-163. https://doi.org/10.1016/j.shpsc.2011.05.014

Nicholson D (2013) “Organisms ≠ Machines.” Studies in History and Philosophy of Biological and Biomedical Sciences 44: 669–678. https://doi.org/10.1016/j.shpsc.2013.05.014

Nicholson D (2014) “The machine conception of the organism in development and evolution: a critical analysis.” Studies in History and Philosophy of Biological and Biomedical Sciences 48: 162-74. https://doi.org/10.1016/j.shpsc.2014.08.003

Nicholson D (2018) “Reconceptualising the Organism: From Complex Machine to Flowing Stream” In Nicholson D, Dupre J (eds.) Everything Flows: Towards a Processual Philosophy of Biology. Oxford University Press, Oxford, pp 139-166

Nijhawan D, Honarpour N, Wang X (2000) “Apoptosis in neural development and disease.” Annual Review of Neuroscience 23: 73-87. https://doi.org/10.1146/annurev.neuro.23.1.73

Novick R, Geisinger E (2008) “Quorum Sensing in Staphylococci.” Annual review of genetics 42: 541-64. https://doi.org/10.1146/annurev.genet.42.110807.091640

O’Toole, Kaplan H, Kolter R (2000) “Biofilm Formation as Microbial Development.” Annual Review of Microbiology 54: 49-79. https://doi.org/10.1146/annurev.micro.54.1.49

Odling-Smee J, Laland K, Feldman M (2003) Niche construction: the neglected process in evolution. Princeton University Press, Princeton

Odling-Smee FJ, Erwin D, Palkovacs E, (2013) “Niche construction theory: a practical guide for ecologists.” Quarterly Review of Biology 88: 3–28. https://doi.org.10.1086/669266

Opferman JT, Korsmeyer SJ (2003) “Apoptosis in the development and maintenance of the immune system.” Nature Immunology 4: 410-5. https://doi.org/10.1038/ni0503-410

Pittendrigh, CS (1958) “Adaptation, Natural Selection and Behavior.” In Roe A, Simpson GG (eds.) Behavior and Evolution. Yale University Press, New Haven, pp 390–419

Potten C, Wilson J (2004) Apoptosis: The Life and Death of Cells. Cambridge University Press, Cambridge

Rattcliffe M (2001) “A Kantian Stance on the Intentional Stance.” Biology and Philosophy 16: 29-52. https://doi.org10.1023/A:1006710821443

Ratzsch D & Koperski J (2019) “Teleological Arguments for God’s Existence.” In Zalta EN (ed.) The Stanford Encyclopedia of Philosophy https://plato.stanford.edu/archives/sum2019/entries/teleological-arguments

Reed JC (2000) “Mechanisms of Apoptosis.” The American Journal of Pathology 157(5): 1415–1430. https://doi.org/10.1016/S0002-9440(10)64779-7

Rees D, Glenner H (2014) “Control region sequences indicate that multiple externae represent multiple infections by Sacculina carcini (Cirripedia: Rhizocephala).” Ecology and Evolution 4: 3290–3297. https:doi.org/10.1002/ece3.1177.

Renehan AG, Booth C, Potten CS (2001) “What is apoptosis, and why is it important?” BMJ 322:1536. https://doi.org/10.1136/bmj.322.7301.1536

Rosenblueth A, Wiener N, Bigelow J (1943) “Behavior, Purpose and Teleology.” Philosophy of Science, 10(1): 18–24. https://doi.org/10.1086/286788

Riskin J (2016) The Restless Clock. The University of Chicago Press, Chicago.

Roy H, Steinkraus C, Eilenberg J, Hajek AE, Pell JK (2006) “Bizarre interactions and endgames: entomopathogenic fungi and their arthropod hosts.” Annual Review of Entomology 51: 331– 357. https://doi.org/10.1146/annurev.ento.51.110104.150941

Rudwick MJS (1964) “The Inference of Function from Structure in Fossils.” The British Journal for the Philosophy of Science XV(57): 27–40, https://doi.org/10.1093/bjps/XV.57.27

Ruiz-Mirazo K, Pereto J, & Moreno A (2004) “A Universal Definition of Life: Autonomy and Open-Ended Evolution.” Origins of life and evolution of the biosphere: the Journal of the International Society for the Study of the Origin of Life 34: 323-46. https://doi.org10.1023/B:ORIG.0000016440.53346.dc.

Ruiz-Mirazo K, Pereto J & Moreno A, (2010) “Defining Life or Bringing Biology to Life.” Origins of life and evolution of the biosphere: the Journal of the International Society for the Study of the Origin of Life 40: 203-13. https://doi.org/10.1007/s11084-010-9201-6.

Ruiz-Mirazo K, Moreno A (2012) “Autonomy in evolution: from minimal to complex life.” Synthese 185(1): 21–52. https://doi.org/10.1007/s11229-011-9874-z

Ruse M (2000) “Teleology: Yesterday, Today and Tomorrow”. Studies in History & Philosophy of Biological and Biomedical Sciences 31: 213-232. https://doi.org/10.1016/S1369-8486(99)00046-1

Saborido C, Mossio M, Moreno A (2011) “Biological organization and cross-generation functions.” British Journal for the Philosophy of Science 62: 583-606. https://doi.org/10.1093/bjps/axq034

Saborido C, Moreno A, González-Moreno M, Clemente J (2016) “Organizational Malfunctions and the Notions of Health and Disease.” In Giroux E (ed.) Naturalism in the Philosophy of Health: Issues and Implications. Springer, Dordrecht, pp 101-120

Sakahira H, Enari M, Nagata S (1998) “Cleavage of CAD inhibitor in CAD activation and DNA degradation during apoptosis.” Nature 391: 96–99. https://doi.org/10.1038/34214

Schiefsky M (2007) “Galen’s Teleology and Functional Explanation”. In Sedley D (ed.) Oxford Studies in Ancient Philosophy. Oxford University Press, Oxford, pp 369–400

Schlosser G (1998) “Self-re-production and functionality: A systems-theoretical approach to teleological explanation.” Synthese 116: 303-54. https://doi.org/10.1023/A:1005073307193

Sloan P (2019) "Evolutionary Thought Before Darwin." In Zalta EN (ed.) The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/win2019/entries/evolution-before-darwin/

Solano C, Echeverz M, Lasa I (2014) “Biofilm dispersion and quorum sensing.” Current Opinion in Microbiology 18C: 96-104. https://doi.org/10.1016/j.mib.2014.02.008.

Stich S (1996) Deconstructing the Mind. Oxford University Press, Oxford.

Stoodley P, Sauer K, Davies DG, Costerton JW (2002) “Biofilm as Complex Differentiated Communities.” Annual Review of Microbiology 56: 187-209. https://doi.org/10.1146/annurev.micro.56.012302.160705.

Thomas F, Adamo S, Moore J (2005) “Parasitic manipulation: where are we and where should we go?” Behavioural Processes 68: 185-199. https://doi.org/10.1016/j.beproc.2004.06.010

Thornberry NA, Lazebnik Y (1998) “Caspases: Enemies Within.” Science 281(5381): 1312-1316. https://doi.org/10.1126/science.281.5381.1312

Turner J (2004) “Extended phenotypes and extended organisms.” Biology and Philosophy 19: 327–352. https://doi.org/10.1023/B:BIPH.0000036115.65522.a1

Vance S (2011) “The effect of the mermithid parasite Gasteromermis sp (Nematoda: Mermithidae) on the drift behaviour of its mayfly host, Baetis bicaudatus (Ephemeroptera: Baetidae): A trade off between avoiding predators and locating food.” Canadian Journal of Zoology 74: 1907-1913. https://doi.org/10.1139/z96-215.

Varela F (1979) Principles of Biological Autonomy. North-Holland, New York.

von Bertalanffy, Ludwig. (1968). General Systems Theory: Foundations, Development, Application. George Braziller Inc, New York.

Watson JD & Crick HC (1953) “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid.” Nature 171: 737-738. https://doi.org/10.1038/171737a0

Weiner N (1948) “Cybernetics: Or Control and Communication in the Animal and the Machine.” The Technology Press, Massachusetts

Wells DA (2015) “The extended phenotype(s): a comparison with niche construction theory.” Biology and Philosophy 30: 547-567. https://doi.org/10.1007/s10539-015-9476-0

West SA, Griffin AS, Gardner A, Diggle SP (2006) “Social evolution theory for microorganisms.” National Review of Microbiology 4: 597–607. https://doi.org/10.1038/nrmicro1461

West SA, Diggle SP, Buckling A, Gardner A, Griffin A (2007) “The Social Lives of Microbes.” Annual Review of Ecology, Evolution, and Systematics 38: 53-77. https://doi.org/10.1146/annurev.ecolsys.38.091206.095740

Wimsatt WC (1972) “Teleology and the Logical Structure of Function Statements.” Studies in History and Philosophy of Science Part A, 3(1): 1–80. https://doi.org/10.1016/0039-3681(72)90014-3

Wright L (1973) “Functions.” The Philosophical Review, 82(2): 139–168. https://doi.org/10.2307/2183766